Linear Algebra
线性代数

Wentao Zhu
Queens' College, University of Cambridge

June 2025

目录

1 Vectors 向量 **7**
1.1 Vector Addition, Scalar Multiplication and Linear Combinations . . . 7
1.2 Euclidean Space $\mathbb{R}^n$ and Hyperplane . . . 7
1.3 Visualization of Vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ . . . 8
1.4 Lengths and Dot Products . . . 9
1.5 Lines in $\mathbb{R}^2$ . . . 10
1.6 Planes and Lines in $\mathbb{R}^3$ . . . 12
1.7 m-Dimensional Geometric Object in $\mathbb{R}^n$ . . . 14
1.8 Exercises . . . 17

2 Matrices 矩阵 **18**
2.1 Definitions, Notation and Terminology . . . 18
2.2 Operations on Matrices . . . 20
2.3 The Laws of Matrix Algebra . . . 21
2.4 The Inverse Matrix and its Properties . . . 21
2.5 Powers of a Matrix . . . 22
2.6 Properties of the Transpose of a Matrix . . . 22
2.7 Block Matrix . . . 23
2.8 Exercises . . . 25

3 Vectors and Matrices 向量与矩阵 **26**
3.1 Revisit Vectors: a special type of Matrix . . . 26
3.2 Revisit Matrices: a set of Vectors . . . 26
3.3 Vector times Vector: 2 Ways (revisit dot product and introduce outer product) . . . 26
3.4 Matrix times Vector and Vector times Matrix: 2 Ways . . . 27
3.5 Matrix times Matrix: 4 Ways . . . 27
3.6 Revisit Scalars: a 1×1 matrix . . . 28
3.7 Inner, Dot, Scalar, Cross, Vector, Outer, Tensor Product . . . 28
3.8 Exercises . . . 30

4 Systems of Linear Equations 线性方程组 31
4.1 Determinant . . . 31
4.2 Properties of Determinant . . . 32
4.3 Calculation of Determinant . . . 33
4.4 The Inverse Matrix and Adjoint Matrix . . . 33
4.5 Properties of the Adjoint Matrix . . . 33
4.6 Cramer's Rule . . . 33
4.7 The Gauss-Jordan Elimination Method . . . 34
4.8 Elementary Operations and Matrix Equivalence . . . 35
4.9 LU/LDU Decomposition . . . 38
4.10 Rank of Matrix . . . 40
4.11 Exercises . . . 42

5 Vector Spaces 向量空间 43
5.1 Vector Space and Subspace . . . 43
5.2 Linear Span and Linear Independence . . . 43
5.3 Basis and Dimension . . . 44
5.4 Normed Vector Space and Inner Product Space . . . 44
5.5 Hilbert Space . . . 46
5.6 Coordinates with respect to a basis . . . 46
5.7 Orthogonality between Subspaces . . . 50
5.8 Sum and Intersection of Subspaces . . . 50
5.9 Complement and Orthogonal Complement of a Subspace . . . 51
5.10 Exercises . . . 53

6 Vector Spaces in Matrices 矩阵中的向量空间 54
6.1 Column Space, Row Space, Null Space . . . 54
6.2 The Rank-Nullity Theorem . . . 55
6.3 The Structure of the Solutions of Linear Systems . . . 55
6.4 Orthogonality between $N(\mathbf{A})$ and $RS(\mathbf{A})$ . . . 55
6.5 Exercises . . . 57

7 Abstract Vector Spaces 抽象向量空间 58
7.1 Revisit Vector Spaces . . . 58
7.2 Examples of Abstract Vector Spaces . . . 59
7.3 Basis in Abstract Vector Spaces and Linear Isomorphism . . . 60
7.4 Other Norms of $\mathbb{R}^n$ . . . 62
7.5 Norms and Inner Products of Abstract Vector Spaces . . . 63
7.6 Exercises . . . 68

8 Linear Transformations I 线性变换 I（基本概念） 69
8.1 Linear Transformations: functions between vector spaces . . . 69
8.2 Linear Operator . . . 69
8.3 Linear VS non-Linear . . . 70

8.4 Matrix as a Linear Transformation 71
8.5 Geometric Interpretation of Linear Transformations 72
8.6 Revisit Inverse, Determinant, Rank and Transpose 74
8.7 Rank-1 Matrix 76
8.8 Low-rank Matrix 77
8.9 Reflections, Rotations and Stretches in $\mathbb{R}^2$ 78
8.10 Orthogonal Matrices: Rigid Transformations 79
8.11 Reflections and Rotations in $\mathbb{R}^3$ 81
8.12 Exercises 82

9 Linear Transformations II 线性变换 II（从标准基推广至任意基） 83

9.1 The Matrix $A_T^{B\to C}$ 83
9.2 Change of Basis and Transition Matrix 84
9.3 The Transition Matrix $P_{B\to B'}$ 85
9.4 Change of Basis and Linear Transformations 88
9.5 Similar Matrix 89
9.6 Diagonalisable Linear Transformations 90
9.7 Eigenvalues, Eigenvectors and Eigenspaces 90
9.8 Relationship between Diagonalisation and Eigenvectors 92
9.9 Conditions of Diagonalisation 92
9.10 Orthogonal Diagonalisation 94
9.11 Symmetric Matrices and Quadratic Forms 95
9.12 Trace and other Similarity Invariants 97
9.13 Exercises 100

10 Linear Transformations III 线性变换 III（关注一类特殊的变换：投影） 101

10.1 Projections 101
10.2 Projections onto Subspaces 101
10.3 Idempotent Matrix 102
10.4 Parallel Projections 105
10.5 Orthogonal Projections 111
10.6 Application of Orthogonal Projection: OLS 114
10.7 Orthogonal Projections in Abstract Vector Spaces 115
10.8 Exercises 118

11 Linear Transformations IV 线性变换 IV（从 $\mathbb{R}$ 推广至 $\mathbb{C}$） 119

11.1 Complex Vector Spaces 119
11.2 Special Complex Matrices 120
11.3 Unitary Diagonalisation 121
11.4 Spectral Decompositions 123
11.5 Exercises 127

12 Linear Transformations V 线性变换 V（从方阵推广至任意矩阵） **128**
12.1 Singular Values . . . 128
12.2 Singular Value Decomposition (SVD) . . . 129
12.3 Eckart–Young Theorem and Low-rank Approximation . . . 130
12.4 Left and Right Inverses . . . 132
12.5 Generalised Inverses . . . 132
12.6 Exercises . . . 134

13 Linear Transformations VI 线性变换 VI（从对角矩阵推广至 JNF） **135**
13.1 The Jordan Normal Form (JNF) . . . 135
13.2 Application Scenario 1: Solving Systems of Difference Equations . . . 137
13.3 Application Scenario 2: Solving Systems of Differential Equations . . . 137
13.4 Application Scenario 3: Dominant Eigenvalues and Long-Term Behaviour . . . 137
13.5 Exercises . . . 138

14 Matrix Algebra 矩阵代数 **139**
14.1 Review: Basics . . . 139
14.2 Derivatives of a Determinant . . . 140
14.2.1 General form . . . 140
14.2.2 Linear forms . . . 140
14.2.3 Square forms . . . 140
14.2.4 Other nonlinear forms . . . 140
14.3 Derivatives of an Inverse . . . 141
14.4 Derivatives of Eigenvalues . . . 141
14.5 Derivatives of Matrices, Vectors and Scalar Forms . . . 141
14.5.1 First Order . . . 141
14.5.2 Second Order . . . 142
14.5.3 Higher-order and non-linear . . . 142
14.5.4 Gradient and Hessian . . . 143
14.6 Derivatives of Traces . . . 143
14.6.1 First Order . . . 143
14.6.2 Second Order . . . 143
14.6.3 Higher Order . . . 144
14.6.4 Other . . . 144
14.7 Derivatives of vector norms . . . 145
14.8 Matrix Decompositions . . . 145
14.8.1 LU Decomposition . . . 145
14.8.2 QR Decomposition . . . 146
14.8.3 Eigen Decomposition (Diagonalisation) . . . 146
14.8.4 Singular Value Decomposition . . . 147
14.8.5 Cholesky Decomposition . . . 147
14.8.6 Schur Decomposition . . . 147
14.8.7 Spectral Decomposition . . . 148

14.8.8 Polar Decomposition 148
14.9 Exercises 149

15 Applications 应用 **150**
15.1 Applications in Statistics 150
15.1.1 Covariance Matrix and Correlation Matrix 150
15.1.2 Principal Component Analysis (PCA) 153
15.2 Applications in Machine Learning 158
15.2.1 Sparse Matrix 158
15.3 Exercises 160

Preface 前言

“There is hardly any theory which is more elementary than linear algebra, in spite of the fact that generations of professors and textbook writers have obscured its simplicity by preposterous calculations with matrices.” —Jean Dieudonné

Reference 参考材料

[1] 线性代数基础与解法全集，一高数，B 站视频
[2] LSE MA100 Lecture Notes
[3] LSE MA212 slides
[4] Cambridge E300 slides
[5] The Art of Linear Algebra, Graphic Notes on “Linear Algebra for Everyone” pdf.
[6] Introduction to Linear Algebra fourth edition, Gilbert Strang
[7] Gilbert Strang Linear Algebra MIT open lectures videos
[8] 高等代数学习指导书，丘维声
[9] 矩阵的三重身份，线代不抽象，B 站视频
[10] 线性代数的几何意义，任广千
[11] 鸢尾花书矩阵力量
[12] Essence of Linear Algebra, 3Blue1Brown, youtube videos

1 Vectors 向量

1.1 Vector Addition, Scalar Multiplication and Linear Combinations

一对独立的数字 v_1 和 v_2 构成一个二维列向量 $\mathbf{v}$：

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

向量加法和标量乘法是向量的两种基本运算。

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \quad \mathbf{v} + \mathbf{w} = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \end{bmatrix}$$

$$c\mathbf{v} = \begin{bmatrix} cv_1 \\ cv_2 \end{bmatrix} \quad -\mathbf{v} = \begin{bmatrix} -v_1 \\ -v_2 \end{bmatrix}$$

数字 c 被称为“标量”。注意，$-\mathbf{v}$ 和 $\mathbf{v}$ 的和是零向量 $\mathbf{0}$，这与数字零不同。

$$\mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

线性代数建立在这两种运算之上——向量加法和标量乘法。将加法与标量乘法结合，我们现在可以构成 $\mathbf{v}$ 和 $\mathbf{w}$ 的“线性组合”：$c\mathbf{v} + d\mathbf{w}$。

向量当然也可以存在于三维以及任意更高维度。

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

我们只关注列向量。行向量仅作为工具或一种表示法出现。下次当我们说“向量”时，我们指的是“列向量”。

1.2 Euclidean Space $\mathbb{R}^n$ and Hyperplane

在我们将向量放入欧几里得空间 $\mathbb{R}^2$ 和 $\mathbb{R}^3$ 以获得几何直观之前，首先正式介绍什么是欧几里得空间 $\mathbb{R}^n$，并阐明我们可能不太直观熟悉的高维欧几里得空间（n>3），将是恰当的。

欧几里得空间（Euclidean space）是一种数学空间，它将熟悉的 2D 和 3D 空间推广到任意有限数量的维度。它记作 $\mathbb{R}^n$，其中 n 代表维数。$\mathbb{R}^n$ 中的每个点由一个有序元组 $(x_1, x_2, ..., x_n)$ 表示，其中每个 x_i 是一个实数。

在之后的学习中我们会了解到，欧几里得空间是有限维的、实的**希尔伯特空间(Hilbert space)**，这意味着它是一个**完备**的空间，并且它是一个**内积空间（inner product space）**。欧几里得空间装备的内积（inner product）为**点积（dot product）**。欧几里得空间是一个内积空间意味着它是一个**赋范向量空间（normed vector space）**。一方面，这进一步意味着它是一个**向量空间（vector space）**；另一方面，这意味着它是一个**度量空间（metric space）**，或者更本质的，它是一个**拓扑空间（topological space）**。相信度量空间的含义我们已经在数学分析和实分析中学到，其余数学空间的概念我们之后会逐一讲解。

由于欧几里得空间是一个度量空间，我们先给出欧几里得空间中两个点之间的**距离函数（distance function）**：两点 $p = (p_1, ..., p_n)$ 和 $q = (q_1, ..., q_n)$ 之间的距离由下式给出：

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$$

欧几里得空间还是一个赋范空间和内积空间，因此它装备了对向量长度的度量（范数）以及内积运算。内积不仅能自然给出范数的定义，还能度量两个向量之间的角度。欧几里得空间中的内积叫点积，我们将在之后详细学习。

$\mathbb{R}^1$ 到 $\mathbb{R}^3$ 可能是最熟悉的，因为很容易在现实世界中可视化它们。$\mathbb{R}^1$ 是一条 1D 直线；$\mathbb{R}^2$ 是一个 2D 平面；$\mathbb{R}^3$ 是我们在现实中体验的熟悉的 3D 空间。对于 $\mathbb{R}^n$ (n>3)，它是一个我们可以进行数学分析但难以可视化的高维空间。

二维平面需要能够找到两条相互正交的向量，三维空间需要能够找到三条相互正交的向量，按这个逻辑，四条相互正交的向量是个什么画面，生活在 3 维世界中的我们似乎难以想象。数学家有一种可视化四维及以上高维空间的方式，或者说简单技巧：把 3 维空间看成一个平面，将垂直于这个平面的方向看作第四维度。同理，n 维空间也可以把 n-1 维空间看成一个平面，我们把这个平面叫做超平面（hyperplane），n 维空间可以被无数个 n-1 维的超平面“切割”或者说“填充”。超平面是 n 维欧几里得空间中的一个平坦的 (n-1) 维子空间。在 $\mathbb{R}^2$ 中，超平面是一条直线，它是一个切割 2D 平面的 1D 对象。在 $\mathbb{R}^3$ 中，超平面是一个平面，它是一个切割 3D 空间的 2D 曲面。在 $\mathbb{R}^n$ 中，超平面总是一个 (n-1) 维的平坦曲面。一个超平面总是将空间分割成两个半空间。

1.3 Visualization of Vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$

可视化能帮助我们更好地理解向量及其运算。让我们从二维情形开始，其中一个向量可以用 xy 平面上的一个箭头表示。下图展示了我们如何可视化向量。基本上，我们可以将向量 $\mathbf{v}$ 描述为两个数字、一个从 (0,0) 出发的箭头或者平面中的一个点。

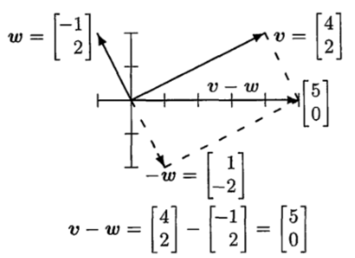

图 1: 二维向量

当我们考虑三维空间中的向量时，xy 平面被三维空间所取代。类似地，列向量与从原点出发的箭头以及箭头终点所指向的点之间存在完美的对应关系。

对于一个向量 $\mathbf{u}$，其唯一的线性组合是标量倍数 $c\mathbf{u}$。对于两个向量，组合是 $c\mathbf{u} + d\mathbf{v}$。对于三个向量，组合是 $c\mathbf{u} + d\mathbf{v} + e\mathbf{w}$。假设向量 $\mathbf{u}, \mathbf{v}, \mathbf{w}$ 在三维空间中，我们已经知道如何在图中找到给定 c, d, e 对应的组合。现在让我们思考一下，所有组合 $c\mathbf{u}$、$c\mathbf{u} + d\mathbf{v}$、$c\mathbf{u} + d\mathbf{v} + e\mathbf{w}$ 构成的图形分别是什么？

答案当然取决于特定的向量 $\mathbf{u}$、$\mathbf{v}$ 和 $\mathbf{w}$。如果它们都是零向量，那么所有的组合都将是零。让我们假设它们都是典型的非零向量。组合 $c\mathbf{u}$ 填满一条无限长的直线。只要 $\mathbf{v}$ 不是 $\mathbf{u}$ 的标量倍数 $c\mathbf{u}$，向量 $\mathbf{u}$ 和 $\mathbf{v}$ 的线性组合就填满一个平面。只要 $\mathbf{w}$ 不在 $\mathbf{u}$ 和 $\mathbf{v}$ 张成的平面上（即 $\mathbf{w}$ 不是 $c\mathbf{u} + d\mathbf{v}$ 的

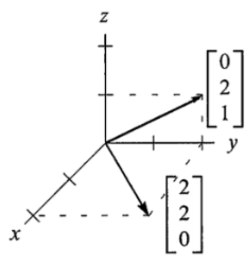

图 2: 三维向量

形式)，那么 $\mathbf{u}$、$\mathbf{v}$ 和 $\mathbf{w}$ 的线性组合就填满整个三维空间。如果 $\mathbf{w}$ 恰好是 $c\mathbf{u}+d\mathbf{v}$ 的形式，即第三个向量位于前两个向量所确定的平面内，那么 $\mathbf{u}$、$\mathbf{v}$ 和 $\mathbf{w}$ 的线性组合将不会超出那个 $\mathbf{uv}$ 平面。

可以初步意识到的是，对于在 $\mathbb{R}^3$ 中的两个向量 $\mathbf{u}$ 和 $\mathbf{v}$，如果 $\mathbf{v}$ 恰好是 $\mathbf{u}$ 的标量倍数，那么在形成线性组合集的工作上，有一个向量是被" 浪费" 的，即没有发挥额外作用的，因为这两个向量的线性组合集是完全一样的，或者说它们互在彼此的线性组合集里。同理，对于在 $\mathbb{R}^3$ 中的三个向量，如果其中一个是另外两个的某种线性组合，那么这个向量在形成线性组合集的任务上也是没用的，因为它已经在另外两个向量的线性组合集里了。也就是说，在 $\mathbb{R}^3$ 中，虽然说"1 个向量的线性组合是直线，2 个向量的线性组合是平面，3 个向量的线性组合是空间" 这种说法是不严谨的，但如果我们默认这 1 个、2 个、3 个向量都是" 有效的"（即线性无关的），那么这样的说法就是成立的。如果给你 3 个非零向量，但它们互为彼此的标量倍数，那么其实我只给了你 1 个有效的向量，则它们的线性组合集是一条直线。

1.4 Lengths and Dot Products

前面我们说过，$\mathbb{R}^n$ 是一个内积空间，这意味着我们无需额外装备，它本身就自带一种内积运算，叫点积。在之后正式介绍向量空间的时候我们会正式给出内积的定义，这一节我们先来简单认识一下点积作为一种内积的运算方式。

在 $\mathbb{R}^n$ 中，向量 $\mathbf{v}$ 和 $\mathbf{w}$ 的**点积**或**内积**是一个数 $\mathbf{v}\cdot\mathbf{w}$：

$$\mathbf{v}\cdot\mathbf{w}=\sum v_i w_i$$

内积 $\mathbf{v}\cdot\mathbf{v}$ 给出了 $\mathbf{v}$ 长度的平方。换句话说，向量 $\mathbf{v}$ 的长度 $\|\mathbf{v}\|$ 是 $\mathbf{v}\cdot\mathbf{v}$ 的平方根：

$$\text{Length}=\text{norm}(\mathbf{v})=\|\mathbf{v}\|=\sqrt{\mathbf{v}\cdot\mathbf{v}}=\sqrt{\sum v_i^2}$$

单位向量 $\mathbf{u}$ 是长度等于 1 的向量，即 $\mathbf{u}\cdot\mathbf{u}=1$。

$\mathbf{u}=\frac{\mathbf{v}}{\|\mathbf{v}\|}$ 是一个与 $\mathbf{v}$ 方向相同的单位向量。

当 $\mathbf{v}$ 垂直于 $\mathbf{w}$ 时，点积 $\mathbf{v}\cdot\mathbf{w}=0$。零向量与每个向量都垂直。

证明：当 $\mathbf{v}$ 和 $\mathbf{w}$ 垂直时，它们构成一个直角三角形的两条边。第三条边是 $\mathbf{v}-\mathbf{w}$。毕达哥拉斯定理告诉我们：

$$\|\mathbf{v}\|^2+\|\mathbf{w}\|^2=\|\mathbf{v}-\mathbf{w}\|^2$$

$$\sum v_i^2+\sum w_i^2=\sum(v_i-w_i)^2$$

$$\sum v_i^2+\sum w_i^2=\sum(v_i^2+w_i^2-2v_iw_i)=\sum v_i^2+\sum w_i^2-2\sum v_iw_i$$

$$\sum v_iw_i=0$$

$$\mathbf{v} \cdot \mathbf{w} = 0$$

现在假设 $\mathbf{v} \cdot \mathbf{w}$ 不为零。$\mathbf{v} \cdot \mathbf{w}$ 的符号立即告诉我们夹角是小于还是大于直角。当 $\mathbf{v} \cdot \mathbf{w}$ 为正时，夹角是锐角。当 $\mathbf{v} \cdot \mathbf{w}$ 为负时，夹角是钝角。

内积揭示了确切的角度 θ。如果 $\mathbf{v}$ 和 $\mathbf{w}$ 是非零向量，则

$$\frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\|\|\mathbf{w}\|} = \cos\theta$$

由于 $|\cos\theta|$ 从不超过 1，余弦公式给出了以下重要不等式（柯西-施瓦茨-布尼亚科夫斯基不等式）：

$$|\mathbf{v} \cdot \mathbf{w}| \le \|\mathbf{v}\|\|\mathbf{w}\|$$

并进一步给出以下不等式（三角不等式）：

$$\|\mathbf{v} + \mathbf{w}\| \le \|\mathbf{v}\| + \|\mathbf{w}\|$$

在证明之前，我们应该注意到这个不等式可以很直观地理解。证明时，一个好的起点是先证明：

$$\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + 2\mathbf{v} \cdot \mathbf{w} + \|\mathbf{w}\|^2.$$

$\mathbf{v} \cdot \mathbf{w} = \cos\theta\|\mathbf{v}\|\|\mathbf{w}\|$ 很好地给出了内积的几何意义：两个向量的内积衡量的是其中一个向量在另一个向量方向上的投影长度，再乘以那个向量的长度。显然，$\mathbf{v}$ 在 $\mathbf{w}$ 方向上的投影长度乘以 $\mathbf{w}$ 的长度，等于 $\mathbf{w}$ 在 $\mathbf{v}$ 方向上的投影长度乘以 $\mathbf{v}$ 的长度（两个直角三角形相似）。这样我们就把两个向量的"积"转化为我们熟悉的两个标量的积，从而得到的结果也是一个标量。有了这个直观理解，我们就可以回过头来验证一些结论，例如，当两个向量正交时，一个向量在另一个向量方向上的投影长度显然就为 0 了，继而这两个向量的内积为 0。我们再考虑两个特殊夹角：当夹角为 $0°$ 时，两个向量的内积就等于它们长度的乘积

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\|\|\mathbf{w}\|$$

当夹角为 $180°$ 时，$\mathbf{v} \cdot \mathbf{w} = -\|\mathbf{v}\|\|\mathbf{w}\|$。我们还可以更进一步地，把内积理解为两个向量在相同方向上的分量的相互影响程度。显然当两个向量方向完全相同时影响最大；两个向量正交，没有相互投影时完全不相关。（注意，投影长度是一个带符号的标量，它只是告诉我们一个方向上的量是正还是负，但它本身不是向量。如果我们关注 $\mathbf{v}$ 在 $\mathbf{w}$ 方向上的投影，那么投影长度自然为 $\|\mathbf{v}\|\cos\theta$。当 $0° < \theta < 90°$ 时，$\cos\theta > 0$，投影是正的，表示 $\mathbf{v}$ 在 $\mathbf{w}$ 的方向上有分量。当 $90° < \theta \le 180°$ 时，$\cos\theta < 0$，投影是负的，表示 $\mathbf{v}$ 在 $\mathbf{w}$ 方向上的分量是反向的。）

如果 $\|\mathbf{v}\| = 5$ 且 $\|\mathbf{w}\| = 3$，那么 $\|\mathbf{v} - \mathbf{w}\|$ 的最小值和最大值是多少？$\mathbf{v} \cdot \mathbf{w}$ 的最小值和最大值又是多少？

1.5 Lines in $\mathbb{R}^2$

假设给定 $\mathbb{R}^2$ 中的一个向量 $\mathbf{p}$ 和一个非零向量 $\mathbf{d}$。我们将通过向量 $\mathbf{p}$ 且方向为向量 $\mathbf{d}$ 的直线定义为 $\mathbb{R}^2$ 的一个子集，该子集由所有满足 $\mathbf{r} = \mathbf{p} + t\mathbf{d}$（其中 t 为实参数）的向量 $\mathbf{r}$ 组成。这个最后的方程被称为通过点 $\mathbf{p}$ 且方向为 $\mathbf{d}$ 的直线的**参数方程**。

在 $\mathbb{R}^2$ 中描述一条直线有三种常见方式：(i) 使用直线上的两个不同点；(ii) 使用直线上的一个点和一个与直线平行的向量；(iii) 使用直线上的一个点和一个与直线正交（也称为垂直）的向量。

情况 (i)：给定直线上的两个不同点 (p_1, p_2) 和 (q_1, q_2)，对应的位置向量是 $\mathbf{p} = \binom{p_1}{p_2}$ 和 $\mathbf{q} = \binom{q_1}{q_2}$。为了构造参数方程，我们需要直线上的一个位置向量（可以选择 $\mathbf{p}$ 或 $\mathbf{q}$）和一个与直线平行的向量。

那么，该直线参数化地描述为所有满足 $\mathbf{x} = \mathbf{p} + t(\mathbf{q} - \mathbf{p})$（其中 t 为实参数）的位置向量 $\mathbf{x}$ 的集合。以分量形式表示，我们有：

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} + t\begin{pmatrix} q_1 - p_1 \\ q_2 - p_2 \end{pmatrix}, \quad t \in \mathbb{R}.$$

为了构造笛卡尔方程（即一般式方程），我们需要直线上的一个位置向量（可以选择 $\mathbf{p}$ 或 $\mathbf{q}$）和一个与直线垂直的向量 $\mathbf{n}$。任何这样的向量被称为**法**向量，并且在乘以一个非零标量意义下是唯一的。给定直线平行于 $\binom{q_1-p_1}{q_2-p_2}$，一个法向量 $\mathbf{n}$ 可以通过交换这些分量并将其中一个乘以负一得到：$\mathbf{n} = \binom{q_2-p_2}{-q_1+p_1}$。这个技巧有效是因为 $\binom{a}{b}$ 与 $\binom{b}{-a}$ 的标量积为零。给定这样的法向量 $\mathbf{n}$，该直线是所有满足 $\mathbf{x} - \mathbf{p}$ 与 $\mathbf{n}$ 垂直的位置向量 $\mathbf{x}$ 的集合；即：

$$(\mathbf{x} - \mathbf{p}) \cdot \mathbf{n} = 0$$

$$(x - p_1)(q_2 - p_2) + (y - p_2)(-q_1 + p_1) = 0$$

这提供了该直线的**笛卡尔方程**（一般式方程）。

获得笛卡尔方程（一般式方程）的另一种方法是从参数方程中消去参数 t：

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} + t\begin{pmatrix} q_1 - p_1 \\ q_2 - p_2 \end{pmatrix}.$$

实际上，假设 $q_1 - p_1$ 和 $q_2 - p_2$ 均不为零，参数 t 可以通过两种不同的方式从这个方程组中解出：要么表示为

$$t = \frac{x - p_1}{q_1 - p_1}$$

要么表示为

$$t = \frac{y - p_2}{q_2 - p_2}.$$

令这两个表达式相等，我们得到一个单一的方程；即：

$$\frac{x - p_1}{q_1 - p_1} = \frac{y - p_2}{q_2 - p_2}.$$

情况 (ii)：给定直线上一点 (p_1, p_2) 和一个与直线平行的向量 $\mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$，可以直接得到参数方程：

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} + t\begin{pmatrix} d_1 \\ d_2 \end{pmatrix}, \quad t \in \mathbb{R}.$$

可以通过前面提到的技巧得到一个法向量：

$$\mathbf{n} = \begin{pmatrix} d_2 \\ -d_1 \end{pmatrix}.$$

因此，笛卡尔方程为：

$$\left\langle \begin{pmatrix} x - p_1 \\ y - p_2 \end{pmatrix}, \begin{pmatrix} d_2 \\ -d_1 \end{pmatrix} \right\rangle = 0.$$

或者写为：

$$(x - p_1)d_2 + (y - p_2)(-d_1) = 0.$$

情况 (iii)： 给定直线上一点 (p_1, p_2) 和一个与直线垂直的向量 $\mathbf{n} = \binom{n_1}{n_2}$，可以直接得到笛卡尔方程：

$$\left\langle \begin{pmatrix} x - p_1 \\ y - p_2 \end{pmatrix}, \begin{pmatrix} n_1 \\ n_2 \end{pmatrix} \right\rangle = 0.$$

或者写为：

$$(x - p_1)n_1 + (y - p_2)n_2 = 0.$$

为了得到参数描述，我们需要一个方向向量 $\mathbf{d}$。这可以通过前面提到的技巧找到：

$$\mathbf{d} = \begin{pmatrix} n_2 \\ -n_1 \end{pmatrix}.$$

因此，参数方程为：

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} + t \begin{pmatrix} n_2 \\ -n_1 \end{pmatrix}, \quad t \in \mathbb{R}.$$

1.6 Planes and Lines in $\mathbb{R}^3$

在 $\mathbb{R}^3$ 中存在一种向量运算，在 $\mathbb{R}^2$ 中没有对应物。这就是以下运算：设

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad \text{和} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

为任意两个向量，并设

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

分别为指向 x, y 和 z 轴正方向的单位向量。向量 $\mathbf{a}$ 和 $\mathbf{b}$ 的**向量积**（或**叉积**）是向量 $\mathbf{a} \times \mathbf{b}$，通过计算所谓的”行列式”（将在后面介绍）来定义：

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \begin{pmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{pmatrix}$$

利用行列式的性质可以证明，向量 $\mathbf{a} \times \mathbf{b}$ 与 $\mathbf{a}$ 和 $\mathbf{b}$ 均正交。

在 $\mathbb{R}^3$ 中描述一个平面有三种常见方式：(i) 使用平面上不共线的三个不同点；(ii) 使用平面上的一个点以及两个平行于该平面但彼此不平行的向量；(iii) 使用平面上的一个点以及一个垂直于该平面的向量。

情况 (i)： 给定平面上三个不共线的不同点 (p_1, p_2, p_3)、(q_1, q_2, q_3) 和 (s_1, s_2, s_3)，对应的位置向量为 $\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$、$\mathbf{q} = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix}$ 和 $\mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix}$。为了构造参数方程，我们需要平面上的一个位置向量（可以选择 $\mathbf{p}$、$\mathbf{q}$ 或 $\mathbf{s}$）以及两个平行于该平面的向量。这样的两个向量是：

$$\mathbf{q} - \mathbf{p} = \begin{pmatrix} q_1 - p_1 \\ q_2 - p_2 \\ q_3 - p_3 \end{pmatrix} \quad \text{和} \quad \mathbf{s} - \mathbf{p} = \begin{pmatrix} s_1 - p_1 \\ s_2 - p_2 \\ s_3 - p_3 \end{pmatrix}$$

设 $\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ 为平面上的一般位置向量。那么，该平面参数化地描述为所有满足 $\mathbf{x} = \mathbf{p} + \lambda(\mathbf{q} - \mathbf{p}) + \mu(\mathbf{s} - \mathbf{p})$（其中 λ 和 μ 为实参数）的位置向量 $\mathbf{x}$ 的集合。以分量形式表示，我们有：

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} + \lambda \begin{pmatrix} q_1 - p_1 \\ q_2 - p_2 \\ q_3 - p_3 \end{pmatrix} + \mu \begin{pmatrix} s_1 - p_1 \\ s_2 - p_2 \\ s_3 - p_3 \end{pmatrix}$$

其中 $\lambda \in \mathbb{R}, \mu \in \mathbb{R}$。

为了构造笛卡尔方程（一般式方程），我们需要平面上的一个位置向量（同样，可以取 $\mathbf{p}$、$\mathbf{q}$ 或 $\mathbf{s}$）以及一个向量 $\mathbf{n}$，该向量同时垂直于方向向量 $\mathbf{q} - \mathbf{p}$ 和 $\mathbf{s} - \mathbf{p}$。找到这样的 $\mathbf{n}$ 的最快方法是使用向量 $\mathbf{q} - \mathbf{p}$ 和 $\mathbf{s} - \mathbf{p}$ 的向量积（叉积）。或者，我们可以通过求解方程组 $\langle \mathbf{n}, \mathbf{q} - \mathbf{p} \rangle = 0$ 和 $\langle \mathbf{n}, \mathbf{s} - \mathbf{p} \rangle = 0$ 来找到 $\mathbf{n}$。无论 $\mathbf{n}$ 是如何得到的，相应的笛卡尔方程将该平面描述为所有满足 $\mathbf{x} - \mathbf{p}$ 与 $\mathbf{n}$ 垂直的位置向量 $\mathbf{x}$ 的集合；即：

$$\langle \mathbf{x} - \mathbf{p}, \mathbf{n} \rangle = 0$$

或者写为：

$$(\mathbf{x} - \mathbf{p}) \cdot \mathbf{n} = 0$$

获得平面笛卡尔方程的另一种方法涉及从参数方程中消去参数 λ 和 μ：

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} + \lambda \begin{pmatrix} q_1 - p_1 \\ q_2 - p_2 \\ q_3 - p_3 \end{pmatrix} + \mu \begin{pmatrix} s_1 - p_1 \\ s_2 - p_2 \\ s_3 - p_3 \end{pmatrix}$$

情况 (ii)： 假设给定平面上一点 (p_1, p_2, p_3) 以及两个平行于该平面但彼此不平行的向量 $\mathbf{u}$ 和 $\mathbf{v}$。那么，记该点的位置向量为 $\mathbf{p}$，我们可以直接写出参数方程：$\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$，$s \in \mathbb{R}$，$t \in \mathbb{R}$。为了构造笛卡尔方程，我们通过使用叉积 $\mathbf{u} \times \mathbf{v}$ 或通过求解方程组 $\langle \mathbf{n}, \mathbf{u} \rangle = 0$ 和 $\langle \mathbf{n}, \mathbf{v} \rangle = 0$ 来找到一个法向量 $\mathbf{n}$。随后得到笛卡尔方程：$\langle \mathbf{x} - \mathbf{p}, \mathbf{n} \rangle = 0$。

情况 (iii)： 给定平面上一点 (p_1, p_2, p_3) 和一个垂直于该平面的向量 $\mathbf{n} = \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix}$，记点 (p_1, p_2, p_3) 的位置向量为 $\mathbf{p}$。可以直接写出笛卡尔方程：$\langle \mathbf{x} - \mathbf{p}, \mathbf{n} \rangle = 0$。为了构造参数方程，我们需要两个平行于该平面的向量 $\mathbf{u}$ 和 $\mathbf{v}$。这样的两个向量总是可以从集合 $\left\{ \begin{pmatrix} 0 \\ n_3 \\ -n_2 \end{pmatrix}, \begin{pmatrix} n_3 \\ 0 \\ -n_1 \end{pmatrix}, \begin{pmatrix} n_2 \\ -n_1 \\ 0 \end{pmatrix} \right\}$ 中选取。随后得到参数方程：$\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$，$s \in \mathbb{R}$，$t \in \mathbb{R}$。

在 $\mathbb{R}^3$ 中描述一条直线有两种常见方式：(i) 使用直线上的两个不同点；(ii) 使用直线上的一个点和一个与直线平行的向量。

情况 (i)： 给定直线上的两个不同点 (p_1, p_2, p_3) 和 (q_1, q_2, q_3)，对应的位置向量为 $\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$ 和 $\mathbf{q} = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix}$。对于参数方程，我们需要直线上的一个位置向量和一个平行于直线的向量，例如

$\mathbf{q}-\mathbf{p}=\begin{pmatrix} q_1-p_1 \\ q_2-p_2 \\ q_3-p_3 \end{pmatrix}$。设 $\mathbf{x}=\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ 为直线上的一般位置向量。那么，该直线参数化地描述为所有满足 $\mathbf{x}=\mathbf{p}+\lambda(\mathbf{q}-\mathbf{p})$（其中 λ 为实参数）的位置向量 $\mathbf{x}$ 的集合。以分量形式表示，我们有：

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}+\lambda\begin{pmatrix} q_1-p_1 \\ q_2-p_2 \\ q_3-p_3 \end{pmatrix}, \quad \lambda\in\mathbb{R}.$$

获得该直线笛卡尔方程（一般式方程）最直接的方法是从参数方程中消去参数 λ。

情况 (ii)：假设给定直线上一点 (p_1,p_2,p_3) 和一个与直线平行的向量 $\mathbf{u}$。那么，记点 (p_1,p_2,p_3) 的位置向量为 $\mathbf{p}$，我们可以直接写出参数方程：$\mathbf{x}=\mathbf{p}+s\mathbf{u}$, $s\in\mathbb{R}$。构造笛卡尔方程的最佳方法是从参数方程组中消去参数 s。这总是会得到一个关于变量 x,y 和 z 的由两个线性方程组成的方程组。

1.7 m-Dimensional Geometric Object in $\mathbb{R}^n$

在 $\mathbb{R}^n$ 空间中，我们可以用两种方式来描述一个 m 维的“平面”或“图形”：参数表达式（parametric form）和方程表达式（Cartesian form 或 implicit form）。

参数表达式的核心思想是：我们可以通过几个“方向向量”来生成这个图形上的所有点。假设我们在 $\mathbb{R}^7$ 中有一个 4 维平面，那么这个图形可以通过一个起点（称为“基点”）和 4 个方向向量来描述。我们引入 4 个自由参数（比如 t_1,t_2,t_3,t_4），每一个参数控制一个方向上的移动。这个表达式大致形式如下：

$$\mathbf{x}=\mathbf{p}+t_1\mathbf{v}_1+t_2\mathbf{v}_2+t_3\mathbf{v}_3+t_4\mathbf{v}_4$$

其中 $\mathbf{p}\in\mathbb{R}^7$ 是图形上的一个点，$\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3,\mathbf{v}_4\in\mathbb{R}^7$ 是方向向量，$t_1,t_2,t_3,t_4\in\mathbb{R}$ 是自由参数。自由参数的个数就是图形的维度：4 个自由度意味着我们在图形内部可以自由移动的方向有 4 个。

与参数表达式不同，方程表达式并不是描述“怎么生成”图形上的点，而是描述“一个点属于这个图形需要满足什么条件”。在 $\mathbb{R}^7$ 中，一个 4 维图形需要满足 3 个独立的线性方程，这些方程限制了点的可能位置，把原本可以在整个 $\mathbb{R}^7$ 中任意变化的点“收缩”到了一个只有 4 个自由度的集合里。这些方程的形式通常是：

$$\begin{cases} a_{11}x_1+a_{12}x_2+\cdots+a_{17}x_7=b_1 \\ a_{21}x_1+a_{22}x_2+\cdots+a_{27}x_7=b_2 \\ a_{31}x_1+a_{32}x_2+\cdots+a_{37}x_7=b_3 \end{cases}$$

每一个方程都将整个空间限制在一个超平面中，而 3 个独立的线性方程会将图形限制在这 3 个超平面交集中的一个更小的子空间——这个子空间的维度就是 7-3=4。

一个很重要的直觉是：空间的总自由度是 n，每加一个线性方程，就多了一个线性限制，就会削去一个自由度，把空间“压缩”一维，最终留下的自由度，就是图形的维度。如果我们对一个点施加了 n-m 个独立的线性限制条件，就会把它限制到一个只有 m 个自由度的图形中。所以：

从自由度的角度理解（构造“有”的部分）：参数表达式中自由参数的个数 = 图形的维度 m 空间被压缩的维度的个数 n-m = 方程表达式中独立方程的个数：从压缩空间的角度理解（削去“无”的部分）

可以这样理解这两种表达方式：参数表达式是一种直接构造法，它告诉我们怎么生成图形上的每一个点。方程表达式是一种限制条件法，它列出所有满足特定线性方程组的点。这些点的解集正是我们所关心的图形。

从方程式到参数式就是解方程的过程。当我们已经有了若干线性方程，我们可以把这看作一个线性方程组。我们关心的是这个方程组的所有解，也就是满足这些条件的所有点。之后我们会介绍一种线性方程组的通用解法：把线性方程组的全部信息写成矩阵，再用高斯消元法求解。这也是为什么线性代数课程中常把“方程组的解结构”和“空间中的图形结构”结合讲解的原因，两者本质上是一回事。

相反，从参数式到方程式可以看作是“反向出题”：给定答案的解集，现在来找出哪些线性方程以这些点为解。步骤也很简单，类似解方程的逆过程：假设参数式为

$$\mathbf{x} = \mathbf{p} + s\mathbf{v}_1 + t\mathbf{v}_2$$

设点 (x, y, z) 是由参数生成的点，用显式表达式列出每个坐标：

$$\begin{cases} x = 1 + s \\ y = s + t \\ z = 2 + t \end{cases}$$

将 s, t 消去，得到点之间的等式关系，结果就是图形满足的线性方程。

注意，很容易想到的一点是，能得到给定解集答案的方程组题目显然不是唯一的。首先，对一个线性方程进行 rescaling，即两边同乘一个非零常数，它的解集不会变。在前面章节中提到的 $\mathbb{R}^2$ 中的直线和 $\mathbb{R}^3$ 中的平面里，rescaling 是唯一得到解集不变的不同方程式的方式，因为只有一个方程。然而，对于 $\mathbb{R}^3$ 中的直线而言，把两个方程进行任意线性组合，依然可以得到解集等价的方程。这些方程的等效变换与之后会介绍的矩阵的初等行变换（elementary row operations）是对应的，它们都是保证方程组的解集不变的变换。最后，当然，如果只需要满足方程组的解集是给定参数式的话，往方程组里添加非独立的重复方程自然也是一种写出不唯一的题目的方式。这也提醒了我们，假设解集是 $\mathbb{R}^5$ 的子空间并且一定有解，3 个线性方程未必代表解集一定是 $\mathbb{R}^5$ 中的 2 维平面，因为 3 个线性方程中可能会有重复的。3 个独立的线性方程才保证解集是 $\mathbb{R}^5$ 中的 2 维平面。之后我们会介绍，我们可以用矩阵的高斯消元法来判断方程组中是否有不是独立的重复方程以及有几个。

反过来，从线性方程组求出参数表达式的结果在形式上其实也不是唯一的，因为我们可以自由选择哪些变量做自由参数，并用不同的方式组织表达式。但如果我们使用统一固定的步骤：先用高斯消元法，再把非主元变量（或者叫自由变量）设为自由参数，将主元变量由自由变量表示，那么得到的参数表达式就是规范化的，在逻辑结构上是唯一的（除了参数名和向量项的顺序）。

例题 1：一个三维图形位于 $\mathbb{R}^5$ 中。它经过点 $\mathbf{p} = (0,0,0,0,0)$，并沿着三个方向向量：

$$\mathbf{v}_1 = (1,0,0,1,0), \quad \mathbf{v}_2 = (0,1,0,0,1), \quad \mathbf{v}_3 = (0,0,1,-1,-1)$$

请写出其参数表达式，并找出该图形满足的两个线性方程。

解：参数表达式为：

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = s\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} + u\begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \\ -1 \end{pmatrix} = \begin{pmatrix} s \\ t \\ u \\ s-u \\ t-u \end{pmatrix}$$

由参数表达式可得：

$$x_4 = x_1 - x_3$$

$$x_5 = x_2 - x_3$$

因此，该图形满足的两个线性方程为：

$$x_1 - x_3 - x_4 = 0$$

$$x_2 - x_3 - x_5 = 0$$

例题 2： 在 $\mathbb{R}^4$ 中，考虑满足以下两个方程的所有点组成的图形：

$$x_1 + x_2 + x_3 + x_4 = 0$$

$$x_1 - x_2 + 2x_3 = 0$$

请写出这个图形的参数表达式。

1.8 Exercises

2 Matrices 矩阵

2.1 Definitions, Notation and Terminology

我们定义大小为 $m \times n$ 的**矩阵**（matrix）A 为一个具有 m 行和 n 列的矩形数阵：

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

数阵中的数字称为矩阵的**元素**（entries）。特别地，出现在 A 的第 i 行和第 j 列的元素记为 a_{ij}，并称为 (i, j) 元素。通常用 $(a_{ij})_{m\times n}$（其中 $1 \leq i \leq m$ 且 $1 \leq j \leq n$）来表示一个 $m \times n$ 矩阵 A 。

两个矩阵相等，当且仅当它们具有相同的大小并且它们对应的元素相等。换句话说，给定

$$A = (a_{ij})_{m\times n} \quad 和 \quad B = (b_{ij})_{m\times n},$$

我们说

$$A = B$$

当且仅当对于所有 i, j（其中 $i \in \{1, 2, ..., m\}$ 且 $j \in \{1, 2, ..., n\}$），有 $a_{ij} = b_{ij}$。

大小为 $m \times n$ 的**零矩阵**（zero matrix），记为 $O = (0)_{m\times n}$，其所有元素均为零：

$$O = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

一个**列矩阵**(column matrix), 也称为**列向量**(column vector), 是只有一列的矩阵 $C = (c_{i1})_{m\times 1}$：

$$C = \begin{pmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{m1} \end{pmatrix}$$

类似地, 一个**行矩阵**(row matrix), 也称为**行向量**(row vector), 是只有一行的矩阵 $R = (r_{1i})_{1\times n}$：

$$R = (r_{11} \quad r_{12} \quad \cdots \quad r_{1n}).$$

具有 n 行和 n 列的矩阵 $S = (s_{ij})_{n\times n}$ 称为 n 阶**方阵**（square matrix of order n）：

$$S = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{pmatrix}.$$

n 阶**单位矩阵**（identity matrix），记为 I_n，是如下矩阵：

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

等价地，我们可以将 I_n 写作 $(a_{ij})_{n\times n}$，其中对于所有 $i \in \{1, 2, ..., n\}$，有 $a_{ii} = 1$，并且如果 $i \neq j$，则 $a_{ij} = 0$。

单位矩阵 I_n 和零矩阵 $O = (0)_{n\times n}$ 是所谓的 n 阶**对角矩阵**（diagonal matrix）的例子。这是一个主对角线以外的所有元素都为零的方阵 $D = (d_{ij})_{n\times n}$：

$$D = \begin{pmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn} \end{pmatrix}.$$

使用替代符号，我们有：$D = (d_{ij})_{n\times n}$，且当 $i \neq j$ 时 $d_{ij} = 0$。

对角矩阵所属的一个更广泛的类别是**三角矩阵**（triangular matrices）。三角矩阵有两种：

一个**上三角矩阵**（upper triangular matrix）$U = (u_{ij})_{n\times n}$ 是一个 n 阶方阵，其中主对角线以下的所有元素都为零：

$$U = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{pmatrix}$$

即，$U = (u_{ij})_{n\times n}$，且当 $i > j$ 时 $u_{ij} = 0$。

类似地，一个**下三角矩阵**（lower triangular matrix）$L = (l_{ij})_{n\times n}$ 是一个 n 阶方阵，其中主对角线以上的所有元素都为零：

$$L = \begin{pmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{pmatrix}$$

换句话说，$L = (l_{ij})_{n\times n}$，且当 $i < j$ 时 $l_{ij} = 0$。

注意，一个 n 阶对角矩阵同时是相同阶数的上三角矩阵和下三角矩阵。

给定如下所示的一般矩阵 $A = (a_{ij})_{m\times n}$，

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

A 的**转置**（transpose），记为 A^T，定义为通过交换 A 的行和列得到的 $n \times m$ 矩阵。即，

$$A^T = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}.$$

一个方阵 $A = (a_{ij})_{n\times n}$ 是**对称**的（symmetric），如果它等于其转置：

$$A = A^T.$$

这等价于说对于所有 i, j，有 $a_{ij} = a_{ji}$。

一个矩阵被称为处于**行阶梯形**（row echelon form），仅当它具有以下三个性质：

1. 对于每一个非零行，该行中第一个非零元素等于 1。这样的元素称为**首项 1**（leading 1）。

2. 对于任意两个非零行，较低行中的首项 1 比较高行中的首项 1 更靠右。

3. 全零行位于矩阵的底部。

一个矩阵被称为处于**简化行阶梯形**（reduced row echelon form），仅当它满足前述三个性质，并附加以下第四个性质：

4. 每个包含首项 1 的列，在该列的其他位置都是零。

2.2 Operations on Matrices

定义在矩阵上的三个主要运算是矩阵加法、标量乘法和矩阵乘法。

矩阵加法运算仅适用于相同大小的矩阵。特别地，给定矩阵 $A = (a_{ij})_{m\times n}$ 和 $B = (b_{ij})_{m\times n}$，它们的和 $A + B$ 定义为通过将 A 的元素与 B 的对应元素相加而得到的矩阵。即，

$$A + B = (c_{ij})_{m\times n}$$

其中

$$c_{ij} = a_{ij} + b_{ij}.$$

标量乘法运算适用于任何大小的矩阵：给定一个矩阵 $A = (a_{ij})_{m\times n}$ 和一个实数 $\lambda \in \mathbb{R}$（也称为**标量**（scalar）），矩阵 λA 定义为通过将矩阵 A 的每个元素乘以 λ 而得到的矩阵。即，

$$\lambda A = (b_{ij})_{m\times n}$$

其中

$$b_{ij} = \lambda a_{ij}.$$

矩阵 λA 称为矩阵 A 的一个**标量倍**（scalar multiple）。

矩阵乘法运算由两个矩阵 A 和 B 产生一个乘积矩阵 AB，仅当 A 的列数等于 B 的行数。特别地，给定一个矩阵 $A = (a_{ij})_{m\times r}$ 和一个矩阵 $B = (b_{ij})_{r\times n}$，它们的乘积是矩阵 $AB = (c_{ij})_{m\times n}$，其一般元素 c_{ij} 由以下规则给出：

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{ir}b_{rj}.$$

该规则告诉我们，为了找到 AB 的一般元素 c_{ij}，我们需要选取 A 的第 i 行和 B 的第 j 列，将该行和该列的对应元素相乘，然后将得到的乘积相加。

作为说明，一个 2×4 矩阵 M 和一个 4×3 矩阵 N 的乘积是 2×3 矩阵 MN，如下所示：

$$MN = \begin{pmatrix} a & b & c & d \\ e & f & g & h \end{pmatrix} \begin{pmatrix} A & B & C \\ D & E & F \\ G & H & I \\ J & K & L \end{pmatrix}$$

$$= \begin{pmatrix} aA + bD + cG + dJ & aB + bE + cH + dK & aC + bF + cI + dL \\ eA + fD + gG + hJ & eB + fE + gH + hK & eC + fF + gI + hL \end{pmatrix}$$

2.3 The Laws of Matrix Algebra

矩阵加法和矩阵乘法运算所满足的五条主要定律罗列并命名如下。在本小节呈现的所有定律中，均假定矩阵的大小使得所指示的运算均可执行。

- 矩阵加法满足**交换律**（commutative）：$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$。
- 矩阵加法满足**结合律**（associative）：$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$。
- 矩阵乘法满足**结合律**（associative）：$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$。
- **左分配律**（left distributive law）：$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$。
- **右分配律**（right distributive law）：$(\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{BA} + \mathbf{CA}$。

此外，还有涉及标量乘法的结合律与分配律。其中四条主要的是：

- $\lambda(\mathbf{AB}) = (\lambda\mathbf{A})\mathbf{B} = \mathbf{A}(\lambda\mathbf{B})$，
- $\lambda(\mu\mathbf{A}) = (\lambda\mu)\mathbf{A}$，
- $\lambda(\mathbf{A} + \mathbf{B}) = \lambda\mathbf{A} + \lambda\mathbf{B}$，
- $(\lambda + \mu)\mathbf{A} = \lambda\mathbf{A} + \mu\mathbf{A}$。

最后，是适当大小的零矩阵 $\mathbf{0}$ 所满足的定律：

- $\mathbf{A} + \mathbf{0} = \mathbf{A}$，
- $\mathbf{A} - \mathbf{A} = \mathbf{0}$，
- $\mathbf{0A} = \mathbf{0}$，
- $\mathbf{A0} = \mathbf{0}$，

以及适当大小的单位矩阵 $\mathbf{I}$ 所满足的定律：

- $\mathbf{AI} = \mathbf{A}$，
- $\mathbf{IA} = \mathbf{A}$。

2.4 The Inverse Matrix and its Properties

矩阵的逆运算仅适用于方阵 $A = (a_{ij})_{n \times n}$。然而，并非所有方阵都可逆。我们称一个矩阵 $A = (a_{ij})_{n \times n}$ 是**可逆的**（invertible）或者**非奇异的**（nonsingular），如果存在一个矩阵 $B = (b_{ij})_{n \times n}$ 使得

$$AB = BA = I_n.$$

回顾 I_n 是 $n \times n$ 单位矩阵。如果 B 存在，则它被称为 A 的**逆**（inverse），并记为 A^{-1}。

定理 2.1 如果 A 是一个 $n \times n$ 可逆矩阵，则矩阵 A^{-1} 是唯一的。

不可逆的方阵又被称为**奇异矩阵**（singular matrix）。我们之后会了解到，它是一个“降维”或“信息丢失”的线性变换。

逆矩阵的一些最重要性质罗列如下：
- 如果 A 是一个可逆矩阵，那么根据定义，A^{-1} 存在且满足

$$AA^{-1} = A^{-1}A = I.$$

这个陈述可以重新解释为：A 是 A^{-1} 的逆，即

$$(A^{-1})^{-1} = A.$$

- 如果 A 是一个可逆矩阵，那么对于任意标量 $\lambda \neq 0$，我们有：

$$(\lambda A)^{-1} = \frac{1}{\lambda} A^{-1}.$$

- 如果 A 和 B 是同阶的可逆矩阵，那么：

$$(AB)^{-1} = B^{-1}A^{-1}.$$

2.5 Powers of a Matrix

给定一个方阵 A 和一个自然数 $n \in \mathbb{N}$，我们将 A^n 定义为乘积

$$A^n = \underbrace{AA\ldots A}_{n\text{ 次}}.$$

矩阵的幂遵循与数的幂相似的规则。首先，如果 A 是一个可逆矩阵且 $n \in \mathbb{N}$，那么：

$$(A^n)^{-1} = (A^{-1})^n.$$

此陈述的证明直接源于逆矩阵的定义和矩阵乘法的结合律。
此外，通常的指数法则也成立；即，对于整数 r, s，我们有：

$$A^r A^s = A^{r+s}$$

和

$$(A^r)^s = A^{rs}.$$

2.6 Properties of the Transpose of a Matrix

转置的一些性质总结如下：

$$(A^T)^T = A,$$

$$(\lambda A)^T = \lambda A^T,$$

$$(A + B)^T = A^T + B^T,$$

$$(AB)^T = B^T A^T,$$

以及最后，

$$(A^T)^{-1} = (A^{-1})^T.$$

2.7 Block Matrix

我们之前处理的矩阵元素都是标量。然而，有时将矩阵视为由若干个小矩阵（称为**子块**（sub-matrix）或**块**（block））组成会非常方便。这样得到的矩阵称为**分块矩阵**（block matrix）。分块矩阵的运算规则与普通矩阵类似，只要子块的维度满足相应的运算要求。

设 A 是一个 $m\times n$ 矩阵。我们可以通过画水平线和垂直线将其划分成若干块。例如，一个 2×2 的分块矩阵形式如下：

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

其中 A_{11} 是 $m_1 \times n_1$ 矩阵，A_{12} 是 $m_1 \times n_2$ 矩阵，A_{21} 是 $m_2 \times n_1$ 矩阵，A_{22} 是 $m_2 \times n_2$ 矩阵，且 $m_1 + m_2 = m$，$n_1 + n_2 = n$。

更一般地，矩阵 A 可以被划分为 $r \times s$ 块：

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1s} \\ A_{21} & A_{22} & \cdots & A_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \cdots & A_{rs} \end{pmatrix}$$

其中每个子块 A_{ij} 是一个 $m_i \times n_j$ 矩阵，满足 $\sum_{i=1}^r m_i = m$ 和 $\sum_{j=1}^s n_j = n$。

设 A 和 B 是两个同为 $m \times n$ 的矩阵，并且按照完全相同的方式分块（即对应的子块具有相同的维度）：

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1s} \\ \vdots & \ddots & \vdots \\ A_{r1} & \cdots & A_{rs} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & \cdots & B_{1s} \\ \vdots & \ddots & \vdots \\ B_{r1} & \cdots & B_{rs} \end{pmatrix}$$

那么它们的和 $C = A + B$ 也是一个具有同样分块结构的矩阵，且满足：

$$C = \begin{pmatrix} A_{11} + B_{11} & \cdots & A_{1s} + B_{1s} \\ \vdots & \ddots & \vdots \\ A_{r1} + B_{r1} & \cdots & A_{rs} + B_{rs} \end{pmatrix}$$

即，分块矩阵的加法可以逐块进行。

设 λ 是一个标量，A 是一个分块矩阵。则标量乘法 λA 定义为：

$$\lambda A = \begin{pmatrix} \lambda A_{11} & \cdots & \lambda A_{1s} \\ \vdots & \ddots & \vdots \\ \lambda A_{r1} & \cdots & \lambda A_{rs} \end{pmatrix}$$

即，标量乘法的效果作用于每一个子块。

分块矩阵的乘法规则与普通矩阵类似，只要子块的维度满足矩阵乘法的要求（即前一个矩阵子块的列数等于后一个矩阵子块的行数）。

设 A 是一个 $m \times n$ 矩阵，分块成 $r \times s$ 块：$A = (A_{ik})_{r\times s}$，其中 A_{ik} 是 $m_i \times n_k$ 矩阵。设 B 是一个 $n \times p$ 矩阵，分块成 $s \times t$ 块：$B = (B_{kj})_{s\times t}$，其中 B_{kj} 是 $n_k \times p_j$ 矩阵。(注意：这种划分要求 A 的列划分方式与 B 的行划分方式完全一致，即 A 的列组维度 $n_1, \ldots, n_s$ 与 B 的行组维度 $n_1, \ldots, n_s$ 相同。)

那么，乘积 $C = AB$ 是一个 $m \times p$ 矩阵，可以分块成 $r \times t$ 块：$C = (C_{ij})_{r\times t}$，其中每个子块 C_{ij} 由下式给出：

$$C_{ij} = \sum_{k=1}^{s} A_{ik}B_{kj}, \quad i = 1, \ldots, r, \quad j = 1, \ldots, t$$

这里的求和指的是矩阵加法。也就是说，分块矩阵的乘法，形式上与普通矩阵乘法完全一致，只是将元素替换为子块，将数的乘法替换为矩阵乘法，将数的加法替换为矩阵加法。

示例： 考虑 2×2 分块矩阵的乘法：

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}$$

只要每个出现的矩阵乘法 $A_{ik}B_{kj}$ 都是有定义的（即 A_{ik} 的列数等于 B_{kj} 的行数），并且每个矩阵加法中的矩阵具有相同的维度，该运算就是有效的。

分块矩阵 A 的转置 A^T 不仅要将行列互换，每个子块自身也需要进行转置。设 $A = (A_{ij})_{r\times s}$ 是一个分块矩阵，则其转置为：

$$A^T = \begin{pmatrix} A_{11}^T & A_{21}^T & \cdots & A_{r1}^T \\ A_{12}^T & A_{22}^T & \cdots & A_{r2}^T \\ \vdots & \vdots & \ddots & \vdots \\ A_{1s}^T & A_{2s}^T & \cdots & A_{rs}^T \end{pmatrix}$$

即，整个块结构的行列索引互换，并且每个单独的子块也取其转置。

分块矩阵在处理大型矩阵、推导公式以及理解矩阵结构时非常有用。一些常见的特殊分块矩阵形式包括：

- **块对角矩阵**（Block Diagonal Matrix）：非零子块仅出现在主对角线上。

$$A = \begin{pmatrix} A_{11} & 0 & \cdots & 0 \\ 0 & A_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{rr} \end{pmatrix}$$

  其性质类似于对角矩阵，例如，其行列式（若定义）和逆矩阵（若存在）都相对容易计算。

- **块上三角矩阵**（Block Upper Triangular Matrix）：主对角线以下的子块全为零。

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1r} \\ 0 & A_{22} & \cdots & A_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{rr} \end{pmatrix}$$

- **块下三角矩阵**（Block Lower Triangular Matrix）：主对角线以上的子块全为零。

$$A = \begin{pmatrix} A_{11} & 0 & \cdots & 0 \\ A_{21} & A_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A_{r1} & A_{r2} & \cdots & A_{rr} \end{pmatrix}$$

分块矩阵是线性代数中一个强有力的工具，它能够简化计算，并帮助我们更深入地理解矩阵运算和矩阵结构。

2.8 Exercises

3 Vectors and Matrices 向量与矩阵

3.1 Revisit Vectors: a special type of Matrix

向量可以被视为一种特殊的矩阵。列向量可以被视为一个 $m \times 1$ 矩阵，即只包含一列的矩阵，称为列矩阵。行向量可以被视为一个 $1 \times n$ 矩阵，即只包含一行的矩阵，称为行矩阵。

3.2 Revisit Matrices: a set of Vectors

一个矩阵可以被视为一组向量。一个 $m \times n$ 矩阵可以被视为 n 个列向量（每个包含 m 个数）或 m 个行向量（每个包含 n 个数）。

3.3 Vector times Vector: 2 Ways (revisit dot product and introduce outer product)

我们已经介绍过一种向量乘以向量的方式，即两个向量的点积。你可能已经注意到，点积 $(\mathbf{a} \cdot \mathbf{b})$ 在矩阵语言中等价于 $\mathbf{a}^T\mathbf{b}$，其结果是一个数。因此，下次当我们想要表示 $\mathbf{a}$ 和 $\mathbf{b}$ 的点积时，除了 $\mathbf{a} \cdot \mathbf{b}$，我们也可以使用矩阵乘法符号 $\mathbf{a}^T\mathbf{b}$，因为它们遵循相同的运算规则，但请记住，当我们写 $\mathbf{a}^T\mathbf{b}$ 时，我们是将 $\mathbf{a}$ 和 $\mathbf{b}$ 视为两个矩阵，因此使用的是矩阵世界中的运算规则。注意，由于点积满足交换律，

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a} = \mathbf{a}^T\mathbf{b} = \mathbf{b}^T\mathbf{a}.$$

在内积被发明时，人们可能尚未发展出矩阵的概念（事实上，情况确实如此）。既然我们现在已经定义了矩阵，我们就拥有了更丰富的数学语言。现在让我们重新考虑一个更一般的问题：给定一个 m 维向量 $\mathbf{a}$ $(m > 1)$ 和一个 n 维向量 $\mathbf{b}$ $(n > 1)$，我们如何定义它们的乘积，以探索除了内积之外是否还有新的答案，因为我们已经知道两个向量可以被视为两个矩阵。请记住，两个矩阵相乘的条件是第一个矩阵的列数应等于第二个矩阵的行数。显然，这里的两个向量不能直接相乘，因为它们的维数不匹配，如果我们希望"两个向量的乘积"这一说法有意义，就需要进行额外的定义。思路很简单：我们想看看是否可以根据两个矩阵的乘积来定义两个向量的乘积（但可能不是原始的两个矩阵，因为我们知道那不可行），因为乘法在矩阵世界中已经是一个定义良好的运算。点积的定义给了我们启发。既然点积在矩阵世界中的形式是 $\mathbf{a}^T\mathbf{b}$，我们自然可以想到另一对可以相乘的矩阵：$\mathbf{a}\mathbf{b}^T$。对于 $\mathbf{a}^T\mathbf{b}$，$\mathbf{a}^T$ 是一个 $1 \times m$ 矩阵，而 $\mathbf{b}$ 是一个 $n \times 1$ 矩阵，但要注意确保两个矩阵可以相乘有一个条件：$m = n$。结果将是一个 1×1 矩阵，即一个标量。对于 $\mathbf{a}\mathbf{b}^T$，$\mathbf{a}$ 是一个 $m \times 1$ 矩阵，而 $\mathbf{b}^T$ 是一个 $1 \times n$ 矩阵。结果将是一个 $m \times n$ 矩阵。注意这里我们不需要 m 和 n 相等。事实上，这确实是一种定义两个向量乘积的方式，称为两个向量的**外积**（outer product），又叫**张量积**（tensor product）。我们可以将向量 $\mathbf{a}$ 和 $\mathbf{b}$ 的外积记为 $\mathbf{a} \otimes \mathbf{b}$。与点积不同，外积不满足交换律。之后我们会学到，外积是一个秩 1 矩阵（rank 1 matrix）并且点积是外积的迹（trace）。

注意，定义两个向量的内积完全不需要引入矩阵，因为它简单地就是两个向量逐元素乘积之和。（这是一个标量的运算。）然而，从矩阵的角度来看，这个运算可以等价地视为第一个矩阵的转置乘以第二个矩阵。事实上，在数学史上，内积最初的定义完全是基于向量本身的定义。后来引入矩阵的影响只是提供了描述它的另一种方式（用矩阵语言描述）。相比之下，外积的定义本质上依赖于矩阵的概念。（它是用矩阵语言定义的！）也就是说，在执行外积运算时，我们必须将两个向量都视为矩阵并进行矩阵乘法。外积本质上就是矩阵乘法。

既然我们现在已经介绍了矩阵，为了保持一致性和便利性，从现在开始我们将尝试使用矩阵语言来描述所有内容。

3.4 Matrix times Vector and Vector times Matrix: 2 Ways

当矩阵与向量相乘时，我们当然将向量视为矩阵。由于矩阵乘法不满足交换律，我们分别讨论矩阵乘向量和向量乘矩阵。

考虑矩阵 A 乘向量 $\mathbf{x}$ 的情况，我们假设 $\mathbf{x}$ 是一个 $n \times 1$ 列向量，那么 A 应该是 $m \times n$ 矩阵，结果是一个 $m \times 1$ 矩阵，即一个列向量。根据矩阵乘法的运算规则，该运算过程可以自然地视为 m 次点积运算。或者，如果我们将 A 视为 n 个列向量，将 $\mathbf{x}$ 视为 n 个标量，那么乘积 $A\mathbf{x}$ 就是 A 的列向量的一个线性组合。

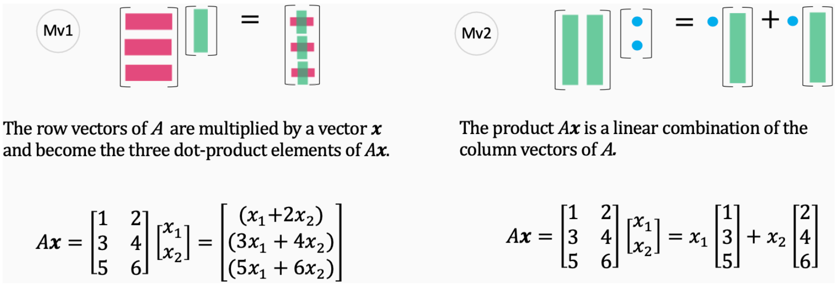

图 3: 矩阵乘向量（2 种视角）

考虑向量 $\mathbf{x}$ 乘矩阵 A 的情况，我们注意到如果 $\mathbf{x}$ 是一个 $n \times 1$ 列向量，那么 A 应该是 $1 \times m$ 矩阵，这实际上是一个向量，因此我们假设 $\mathbf{x}$ 是一个 $1 \times m$ 行向量，那么 A 应该是 $m \times n$ 矩阵，结果是一个 $1 \times n$ 矩阵，即一个行向量。如果我们直接按照矩阵乘法的运算规则来看，该运算过程同样是 n 次点积运算。或者，如果我们将 $\mathbf{x}$ 视为 m 个标量，将 A 视为 m 个行向量，那么乘积 $\mathbf{x}A$ 就是 A 的行向量的一个线性组合。

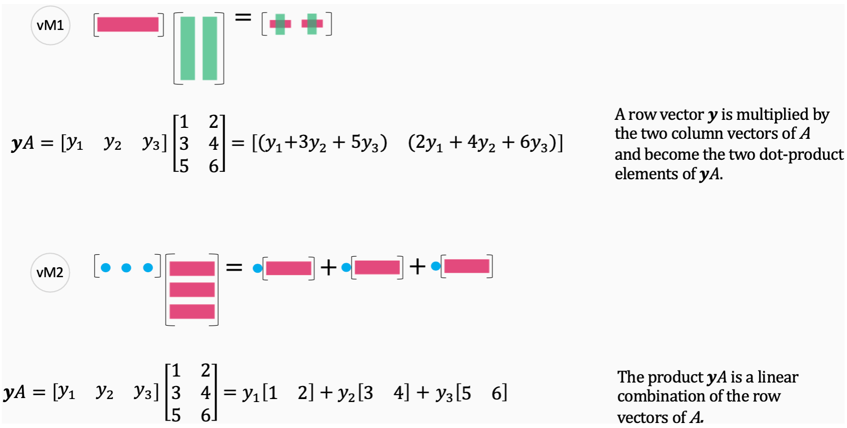

图 4: 向量乘矩阵（2 种视角）

3.5 Matrix times Matrix: 4 Ways

“矩阵乘向量”自然地推广到“矩阵乘矩阵”。让我们考虑一个 $m \times k$ 矩阵 A 乘一个 $k \times n$ 矩阵 B。从运算规则的直接视角来看，每个元素都变成了行向量与列向量的点积。

或者，我们可以通过将 B 视为 n 个列向量，使这种情况变为 n 次“矩阵乘向量”，因此结果矩阵的每一列都是 A 的列向量的线性组合。

或者，我们可以通过将 A 视为 m 个行向量，使这种情况变为 m 次"向量乘矩阵"，因此结果矩阵的每一行都是 B 的行向量的线性组合。

或者，我们可以使这种情况变为 k 次外积，其中结果矩阵是由 A 的列向量与 B 的行向量的外积生成的 k 个矩阵之和。

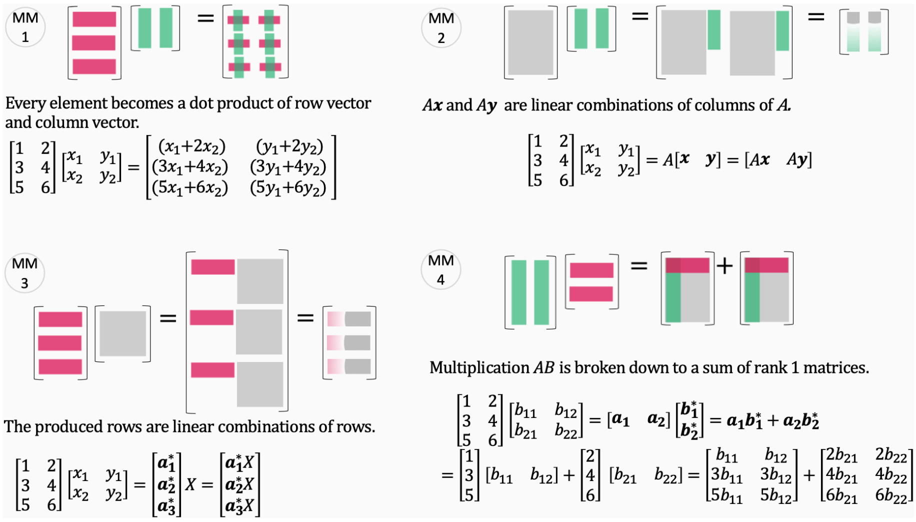

图 5: 矩阵乘矩阵（4 种视角）

3.6 Revisit Scalars: a 1×1 matrix

我们已经说过向量可以看作是行数或列数为 1 的一类特殊的矩阵，这样的统一有助于我们进行矩阵乘法以及数学分析。同样的思想可以更极端地应用在一个矢量上，我们可以把一个矢量看作一个 1×1 的矩阵。一个直接的好处是，当我们处理一个矢量与一个行矩阵或列矩阵相乘时，我们无需定义数乘，而是可以直接将其视为矩阵乘法运算。这带来一个直接的启示：为保证矩阵乘法是合法的，当矢量乘行矩阵（$1 \times n$）时应位于矩阵左边；当矢量乘列矩阵（$m \times 1$）时应位于矩阵右边（"左行右列"）。

3.7 Inner, Dot, Scalar, Cross, Vector, Outer, Tensor Product

现在让我们总结一下内积、点积、标量积、叉积、向量积、外积和张量积之间的区别。

(1) 内积

- **定义**：点积在抽象向量空间中的推广。
- **数学形式**：给定内积空间中的两个向量 $\mathbf{a}, \mathbf{b}$，它们的内积为：

$$\langle \mathbf{a}, \mathbf{b} \rangle$$

- **示例（欧几里得空间）**：在 $\mathbb{R}^n$ 中，内积是点积：

$$\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T \mathbf{b} = \sum a_i b_i$$

- **推广**：在复向量空间 $\mathbb{C}^n$ 中，内积包含共轭：

$$\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a} \cdot \overline{\mathbf{b}} = \mathbf{a}^T \overline{\mathbf{b}} = \sum a_i \overline{b_i}$$

- **直观理解**：衡量向量之间的对齐程度（即投影和角度关系）。

(2) 点积（又叫标量积）
- **定义**：欧几里得空间 $\mathbb{R}^n$ 中的一种特定内积。
- **几何解释**：

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos \theta$$

其中 θ 是向量之间的夹角。
- **结果**：一个标量（数）。
- **性质**：- 交换律：$\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$ - 分配律：$\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$

(3) 外积（又叫张量积）
- **定义**：产生一个矩阵而不是标量。
- **数学形式**：

$$\mathbf{a} \otimes \mathbf{b} = \mathbf{a}\mathbf{b}^T$$

- **性质**：- 不满足交换律。- 秩-1 矩阵（如果 $\mathbf{a}, \mathbf{b}$ 非零）。

(4) 叉积（又叫向量积）
- **定义**：一种结果是另一个向量的乘积（不是标量或矩阵）。
- **数学形式**：

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

展开为：

$$\mathbf{a} \times \mathbf{b} = \begin{bmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{bmatrix}$$

- **仅在 3D 空间中定义**。
- **几何解释**：- 结果向量垂直于 $\mathbf{a}$ 和 $\mathbf{b}$。- 模长：

$$|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}| \sin \theta$$

其中 θ 是向量之间的夹角。
- **性质**：- 不满足交换律：$\mathbf{a} \times \mathbf{b} = -(\mathbf{b} \times \mathbf{a})$。- 分配律：$\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$。

运算	符号	结果	定义域	关键性质
内积	$\langle \mathbf{a}, \mathbf{b} \rangle$	标量	内积空间（如 $\mathbb{R}^n$, $\mathbb{C}^n$, 希尔伯特空间）	点积的推广
点积（标量积）	$\mathbf{a} \cdot \mathbf{b}$	标量	欧几里得空间 $\mathbb{R}^n$	衡量相似性
外积（张量积）	$\mathbf{a} \otimes \mathbf{b}$	矩阵	一般线性代数/矩阵空间（如 $\mathbb{R}^{m \times n}$）	产生秩-1 矩阵
叉积（向量积）	$\mathbf{a} \times \mathbf{b}$	向量	三维欧几里得空间 $\mathbb{R}^3$	垂直于两个输入向量

3.8 Exercises

4 Systems of Linear Equations 线性方程组

4.1 Determinant

考虑小学就学过的多元一次方程组：

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1\\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2\\
a_{31}x_1 + a_{32}x_2 + \cdots + a_{3n}x_n &= b_3\\
&\vdots\\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m
\end{aligned}$$

现在可以被很简单地用矩阵表示：

$$A\mathbf{x} = \mathbf{b}$$

A 是一个系数矩阵，存放了所有方程的系数，这就体现了矩阵的第一个功能：数据存储器。对于求解一个多元一次方程组，或者叫线性系统而言，我们首先想知道它有没有解。我们知道方程组可能会无解，可能会有无穷解，可能会有唯一解。有解我们就称这个系统是**相容的** (consistent)，无解我们就称**不相容** (inconsistent)。

我们希望探索系统有唯一解的情况，所以我们研究 $n \times n$ 的系统，也就是矩阵 A 是正方形矩阵，因为 $m > n$ 时要么有重复的方程从而可以消成 $n \times n$ 的，要么一定无解；$m < n$ 时一定不会有唯一解。对于 $n \times n$ 系统来说，可能有唯一解，即 n 个限制条件既不冲突也没有重复；可能有无穷解，即有重复的限制条件；可能无解，即有矛盾的限制条件。可以看到，如果矩阵 A 可逆，也就是 A^{-1} 存在，那么系统存在唯一解 $\mathbf{x} = A^{-1}\mathbf{b}$。$A\mathbf{x} = \mathbf{b}$ 有唯一解等价于 A^{-1} 存在。

我们可以先从 1×1 开始探索，

$$ax = b$$

x 有唯一解 $x = \frac{b}{a}$ 当且仅当 $a \neq 0$。

同理，对 2×2 我们可以通过消元法算出，$\binom{x_1}{x_2}$ 有唯一解 $\left(\frac{b_1a_{22}-b_2a_{12}}{a_{11}a_{22}-a_{21}a_{12}}, \frac{b_2a_{11}-b_1a_{21}}{a_{11}a_{22}-a_{21}a_{12}}\right)$ 当且仅当 $a_{11}a_{22} - a_{21}a_{12} \neq 0$。

因此，我们可以把这个东西叫做矩阵的行列式，用来判别该矩阵是否可逆。如果行列式不等于 0，则系统有唯一解，矩阵 A 可逆；如果行列式 $= 0$，则系统没有唯一解，矩阵 A 不可逆。

$n \times n$ 矩阵 A 的行列式，记作 $|A|$ 或 $\det A$，是一个从 A 推导出的数字，它决定了矩阵 A 是否可逆。

$$D_1 = |a| = a$$

$$D_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

$$D_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

三阶及高阶行列式有一个计算技巧, 我们首先定义方阵的 minor 和 cofactor: 给定一个 $n\times n$ 矩阵 A, A 的 (i,j) **余子式** (minor), 记作 M_{ij}, 是通过删除 A 的第 i 行和第 j 列后得到的 $(n-1)\times(n-1)$ 矩阵的行列式。

矩阵 A 的 (i,j) **代数余子式** (cofactor)，记作 C_{ij}，由对应的余子式通过以下公式导出：

$$C_{ij} = (-1)^{i+j} M_{ij}$$

有了这些定义，我们可以提供一个计算任何 $n\times n$ 矩阵 A 的行列式的递归公式：选择 A 的任意一行或任意一列，将该行或列中的每个元素乘以 A 的对应代数余子式。然后对这些单独的乘积求和，得到 A 的行列式 $|A|$。

所以你会发现，一个方阵系统有没有唯一解跟常数项 $\mathbf{b}$ 无关，$\mathbf{b}$ 只影响是无解还是无穷解。

总结一下，如果A是一个$n\times n$矩阵，那么以下陈述是等价的：

- $A\mathbf{x} = \mathbf{b}$ 对任何$\mathbf{b}\in\mathbb{R}^n$ 都有唯一解。
- A^{-1} 存在。
- $|A| \neq 0$

4.2 Properties of Determinant

如果 A 是一个 $n\times n$ 矩阵，则 $|A^T| = |A|$。

$$|\lambda A| = \lambda^n |A|$$

$$|ABC| = |A||B||C|$$

如果 $|A| \neq 0$ 的话，

$$|A^{-1}| = \frac{1}{|A|}$$

如果一个 $n\times n$ 矩阵 A 的某一行或某一列全为零，则 $|A| = 0$。

如果将某一行的代数余子式与另一不同行的对应元素相乘并求和，则结果为 0。换句话说，

$$a_{j1}C_{i1} + a_{j2}C_{i2} + \cdots + a_{jn}C_{in} = 0$$

对所有 $i \neq j$ 成立。

如果一个 $n\times n$ 矩阵 A 是对角矩阵、上三角矩阵或下三角矩阵，则 $|A| = a_{11}a_{22}\ldots a_{nn}$。

- 如果交换两行（列），则 $|A|$ 变为 $-|A|$。**推论**：如果一个 $n\times n$ 矩阵 A 有两行或两列相同，则 $|A| = 0$。

- 如果 A 的某一行（列）乘以一个非零常数 α，则 $|A|$ 变为 $\alpha|A|$。**推论**：若某行（列）有公因子 α，则 α 可以提出。若有两行（列）成比例，则行列式值为 0。

- 如果矩阵某行（列）的元素均为两个数之和，则其行列式可以分解为两个行列式之和。**推论**：一次只能拆一行（列），其它元素必须保持不变。（单行可拆性）

$$\begin{vmatrix} a+x & b+y \\ c+z & d+w \end{vmatrix} = \begin{vmatrix} a & b+y \\ c & d+w \end{vmatrix} + \begin{vmatrix} x & b+y \\ z & d+w \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a & y \\ c & w \end{vmatrix} + \begin{vmatrix} x & b \\ z & d \end{vmatrix} + \begin{vmatrix} x & y \\ z & w \end{vmatrix}$$

- 如果将某行的倍数加到另一行上，则 $|A|$ 不变。

4.3 Calculation of Determinant

除了根据定义硬算，有一些技巧：第一种思路就是利用上一节讲的行列式的性质，通过行/列变换把方阵化成行最简形式，这样 RRE(A) 是一个上三角矩阵，行列式就等于对角线上元素的乘积。这个中文教材里叫"打洞法"。

$$D_4 = \begin{vmatrix} 3 & 1 & -1 & 2 \\ -5 & 1 & 3 & -4 \\ 2 & 0 & 1 & -1 \\ 1 & -5 & 3 & -3 \end{vmatrix}$$

4.4 The Inverse Matrix and Adjoint Matrix

我们已经知道了怎么判断一个线性系统是否有唯一解，知道了怎么求一个方阵的行列式，我们接下来讨论如果有唯一解该怎么求解这个唯一解，即怎么求 A 的逆矩阵。

一个方阵 A 的**伴随矩阵** $\mathrm{adj}(A)$ 是它的代数余子式矩阵的转置矩阵。

易证，$A \cdot \mathrm{adj}(A) = \mathrm{adj}(A) \cdot A = |A| I_n$

已知 $|A| \neq 0$，则 $A \cdot \frac{1}{|A|}\mathrm{adj}(A) = I_n$，即 $A^{-1} = \frac{1}{|A|}\mathrm{adj}(A)$

例题： 设 n 阶矩阵 A 满足 $A^2 - 2A - 3I_n = 0$ (1). 证明 A 为可逆矩阵，并求出 A^{-1}。(2). 证明 $(A + 2I_n)$ 为可逆矩阵，并求出 $(A + 2I_n)^{-1}$。

例题： 设 A 为 n 阶非零矩阵，若 $A^3 = 0$，证明：$A - I$ 和 $A + I$ 可逆。

4.5 Properties of the Adjoint Matrix

我们可以来单独研究一下伴随矩阵的性质。首先，最直接的，它等于矩阵的行列式乘上矩阵的逆。(为了简便，我们暂时把 $\mathrm{adj}(A)$ 写作 A^*)

$$A^* = |A| A^{-1}$$

继而可以推得：

$$(kA)^* = k^{n-1} A^*$$

$$|A^*| = |A|^{n-1}$$

$$(A^*)^* = |A|^{n-2} A$$

另外，

$$(AB)^* = B^* A^*$$

4.6 Cramer's Rule

如果 A 是一个 $n \times n$ 可逆矩阵，所谓的克拉默法则为我们提供了求解线性方程组 $A\mathbf{x} = \mathbf{b}$ 的另一种方法。

考虑一个 $n \times n$ 矩阵 A 且 $|A| \neq 0$。令 $\mathbf{x}$ 表示未知向量 $x_1, x_2, ..., x_n$。那么对于任何 $n \times 1$ 列向量 $\mathbf{b}$，线性方程组 $A\mathbf{x} = \mathbf{b}$ 的解由下式给出：

$$x_i = \frac{|A_i|}{|A|}$$

其中 A_i 定义为通过将 A 的第 i 列替换为向量 $\mathbf{b}$ 而从 A 得到的矩阵。

4.7 The Gauss-Jordan Elimination Method

我们知道了如何判断一个线性系统是否有唯一解并且如何求出该唯一解，我们可以先求出 A 的逆矩阵也可以通过克拉默法则来单独求出 $\mathbf{x}$ 向量的每一个分量值。但不管哪种方法，如果不存在唯一解，我们就只能止步于此，对于是否是无解还是无穷解都是无从获知的，并且目前的讨论只局限于 A 是方阵的情况。本节我们就对更一般的 A 为 $m \times n$ 矩阵的情况进行讨论，依然旨在回答两个最基本的问题：如何判断解的情况以及如何求解。

回忆我们小学学过的求解线性方程组的方法：加减消元法：

$$\begin{cases} 4x+2y=94 \\ x+y=35 \end{cases} \to \begin{cases} x+y=35 \\ 4x+2y=94 \end{cases} \to \begin{cases} x+y=35 \\ 2x+y=47 \end{cases} \to \begin{cases} x+y=35 \\ -y=-23 \end{cases} \to \begin{cases} x=12 \\ y=23 \end{cases}$$

矩阵就是一个数表，可以存储方程组的所有信息，我们可以用一个增广矩阵（A 和 $\mathbf{b}$ 构成的分块矩阵）来表示整个系统。刚才的加减消元过程就等价于给这个增广矩阵 $(A|\mathbf{b})$ 做所谓的"初等行变换"来化为 $(\text{RRE}(A)|\mathbf{c})$，我们把这个解线性系统的方法叫做高斯消元法：

$$\left(\begin{array}{cc|c} 4 & 2 & 94 \\ 1 & 1 & 35 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & 1 & 35 \\ 4 & 2 & 94 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & 1 & 35 \\ 2 & 1 & 47 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & 1 & 35 \\ 0 & -1 & -23 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & 0 & 12 \\ 0 & 1 & 23 \end{array}\right)$$

初等行变换 (elementary row operations) 包括 3 类：

(1) 交换两行 (2) 把某一行乘上非零常数 k (3) 把某一行的 k 倍加到另一行上

观察可得，RRE(A) 中 leading one 的个数等于有效方程的个数，我们称为矩阵 A 的秩 (rank)。

行满秩: $\rho(A) = m$

列满秩: $\rho(A) = n$

我们总结一下解的情况的判别：

无解 $\iff \text{rank}(A) < \text{rank}(A|\mathbf{b})$

有解 $\iff \text{rank}(A) = \text{rank}(A|\mathbf{b})$

  唯一解 $\iff \text{rank}(A) = \text{rank}(A|\mathbf{b}) = n$

  无穷解 $\iff \text{rank}(A) = \text{rank}(A|\mathbf{b}) < n$

对所有 $\mathbf{b}$ 都有解 $\iff \text{rank}(A) = \text{rank}(A|\mathbf{b}) = m \iff A$ 是行满秩

对某些 $\mathbf{b}$ 有解，对某些 $\mathbf{b}$ 无解 $\iff \text{rank}(A) < m$

如果已知对某个 $\mathbf{b}$ 有解，A 是列满秩 $\iff$ 解是唯一的

秩的条件	$A\mathbf{x} = \mathbf{b}$ 对所有$\mathbf{b}$ 有解?	如果有解，是否唯一?
A是行满秩($\text{rank}(A) = m$)	对所有$\mathbf{b}$ 都有解	可能有无穷多解（当$n > m$ 时）
A是列满秩($\text{rank}(A) = n$)	可能有些$\mathbf{b}$ 无解	如果有解，则唯一
A是方阵($m = n$)且满秩	对所有$\mathbf{b}$有解	唯一解$\mathbf{x} = A^{-1}\mathbf{b}$

有无解就是看有无矛盾的限制条件，矛盾体现在 $\text{rank}(A) \neq \text{rank}(A|\mathbf{b})$，这意味着最后一行是 $0 =$ 某个常数。正常有解的情况 $\text{rank}(A)$ 都是等于 $\text{rank}(A|\mathbf{b})$ 的，如果已知有解，考虑是否唯一，那就是看有效的限制条件个数是否足够能锁定到只有一个解，$\text{rank}(A)$ 是有效限制条件的个数，等于 n 就是等于未知数的个数，那么 n 个未知数 n 个有效限制条件，就可以求出唯一解。如果 $\text{rank}(A)$ 小于 n，就说明有效限制条件个数小于未知数个数，只能求出无穷解。我们目前的理解只需要通过加减消元的等价过程理解到把矩阵化成 RRE 可以通过 leading ones 的个数快速准确地判断有效方程的个数并且帮助我们求解即可。核心思想其实都是小学学过的，没有变，比如能否求出唯一解要看有效等量关系的个数是否等于未知数的个数，是否会无解取决于是否有矛盾的等量关系。我们目前为止还是纯代数层面的分析，之后我们会对线性系统增补几何层面的理解。

无解和唯一解我们都可以直接写出答案，无穷解要怎么表示？把 leading one 的变量看作主变量，把不是 leading one 的变量看作自由参数，用来表示主变量即可。

下面我们讨论一种特殊情况，当 $\mathbf{b} = \mathbf{0}$ 时，我们称系统为齐次的，齐次线性系统：$A\mathbf{x} = \mathbf{0}$。

一个齐次线性系统是一定有解的，因为不管怎么样 $\mathbf{x} = \mathbf{0}$ 总是它的一个解。当然如果一个齐次线性系统有唯一解，那么也一定是这个平凡解 $\mathbf{x} = \mathbf{0}$。

如果 A 是一个 $m \times n$ 矩阵且 $m < n$，那么 $A\mathbf{x} = \mathbf{0}$ 有无穷多解。因为首先齐次保证有解，$m < n$ 又保证一定不会有唯一解，那么就只能是无穷解了。

最后我们重新审视一下之前的关于方阵的等价结论，扩展一下：

如果A 是一个$n \times n$ 矩阵，那么以下陈述是等价的：

- $A\mathbf{x} = \mathbf{b}$ 对任何$\mathbf{b} \in \mathbb{R}^n$ 都有唯一解。
- A^{-1} 存在。
- $|A| \neq 0$
- $A\mathbf{x} = \mathbf{0}$ 只有平凡解$\mathbf{x} = \mathbf{0}$
- $\text{RRE}(A) = I_n$
- $\text{rank}(A) = n$

4.8 Elementary Operations and Matrix Equivalence

矩阵的**初等变换**（Elementary Operations）是矩阵理论中的基本操作：

1. **初等行变换**（Elementary Row Operations）：
   (a) **互换**（Row Switching）：交换矩阵的两行。
   (b) **倍乘**（Row Multiplication）：用一个非零常数乘矩阵的某一行。
   (c) **倍加**（Row Addition）：将矩阵的某一行的倍数加到另一行上。
2. **初等列变换**（Elementary Column Operations）：
   (a) **互换**（Column Switching）：交换矩阵的两列。
   (b) **倍乘**（Column Multiplication）：用一个非零常数乘矩阵的某一列。
   (c) **倍加**（Column Addition）：将矩阵的某一列的倍数加到另一列上。

基于初等变换，可以定义以下几种矩阵等价关系：

- **矩阵行等价**（Row Equivalence）：如果矩阵 A 可以经过有限次初等**行**变换变成矩阵 B，则称 A 与 B 行等价，记作 $A \overset{r}{\sim} B$。
  意义：行等价的矩阵所对应的线性方程组是同解的。
- **矩阵列等价**（Column Equivalence）：如果矩阵 A 可以经过有限次初等**列**变换变成矩阵 B，则称 A 与 B 列等价，记作 $A \overset{c}{\sim} B$。
  意义：列等价的矩阵其列向量组是等价的（即可以相互线性表出）。
- **矩阵等价**（Matrix Equivalence）：如果矩阵 A 可以经过有限次初等变换（包括行变换和列变换）变成矩阵 B，则称 A 与 B 等价，记作 $A \sim B$。
  意义：两个同型矩阵等价的充要条件是它们的秩相等。

矩阵的等价关系（包括行等价、列等价和一般等价）均满足：

- **反身性**（Reflexivity）：$A \sim A$
- **对称性**（Symmetry）：若 $A \sim B$，则 $B \sim A$
- **传递性**（Transitivity）：若 $A \sim B$ 且 $B \sim C$，则 $A \sim C$

由单位矩阵 I 经过**一次**初等变换得到的矩阵称为**初等矩阵**（Elementary Matrix）。
初等矩阵主要有三种类型，与初等变换一一对应：

1. **互换初等阵**（Permutation Matrix）：由单位矩阵交换第 i 行与第 j 行（或列）得到。

$$E_{ij} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \cdots & 1 & & \\ & & \vdots & \ddots & \vdots & & \\ & & 1 & \cdots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

   例如，交换 2、3 行：$E_{2,3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$

2. **倍乘初等阵**（Scaling Matrix）：由单位矩阵的第 i 行（或列）乘以非零常数 k 得到。

$$E_i(k) = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & k & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}$$

   例如，第 2 行乘 k：$E_2(k) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{pmatrix}$

3. **倍加初等阵**（Row/Column Addition Matrix）：由单位矩阵的第 i 行的 k 倍加到第 j 行上得到（或以列操作方式定义）。

$$E_{ij}(k) = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \cdots & k & & \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

例如，第 1 行的 k 倍加到第 3 行：$E_{1,3}(k) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ k & 0 & 1 \end{pmatrix}$

初等变换可以用矩阵乘法表示。初等矩阵与初等变换的关系可以概括为"**左行右列**"法则：

- 对矩阵 A **左乘**一个初等矩阵 E（即 EA），等价于对 A 施行一次与 E 对应的**初等行变换**。
- 对矩阵 A **右乘**一个初等矩阵 E（即 AE），等价于对 A 施行一次与 E 对应的**初等列变换**。

下面给出初等矩阵的一些性质：

1. 初等矩阵都是**可逆**的。
2. 初等矩阵的逆矩阵仍然是同类型的初等矩阵，即"变换回去"：
   - 互换初等阵的逆是它自身：$E_{ij}^{-1} = E_{ij}$
   - 倍乘初等阵的逆是倍乘其倒数：$E_i(k)^{-1} = E_i(1/k)$
   - 倍加初等阵的逆是倍加其相反数：$E_{ij}(k)^{-1} = E_{ij}(-k)$

对于一个可逆的 $n \times n$ 矩阵 A，对其施加一系列初等行变换可以将其化为简化行阶梯形 $RRE(A)$。根据定理，$RRE(A) = I_n$。设这些行变换对应的初等矩阵分别为 $E_1, E_2, \ldots, E_k$，则有：

$$E_k E_{k-1} \ldots E_2 E_1 A = I_n.$$

令 $P = E_k E_{k-1} \ldots E_2 E_1$，则上式即为 $PA = I_n$。这意味着 P 是 A 的一个左逆。根据可逆矩阵的性质，P 就是 A 的逆矩阵：

$$P = E_k E_{k-1} \ldots E_2 E_1 = A^{-1}.$$

现在，考虑对单位矩阵 I_n 施加同样的初等行变换序列，即计算：

$$E_k E_{k-1} \ldots E_2 E_1 I_n.$$

根据上式，这正好等于 $A^{-1} I_n = A^{-1}$。

由此我们得到一个重要的结论和实用的算法：

- **核心结论**：将可逆矩阵 A 化为单位矩阵 I_n 所使用的那一系列初等行变换，**同时**也会将单位矩阵 I_n 化为 A 的逆矩阵 A^{-1}。

- **求逆算法**：构造一个分块矩阵 $(A \mid I_n)$，然后对其只施行初等**行**变换。当子块 A 被化为 I_n 时，右边的子块 I_n 就同时被化为了 A^{-1}。整个过程可以表示为：

$$(A \mid I_n) \xrightarrow{\text{初等行变换}} (I_n \mid A^{-1})$$

从关系式 $E_k E_{k-1} \dots E_2 E_1 = A^{-1}$ 出发，我们可以自然地得出一个深刻的结论。对等式两边同时取逆，并利用初等矩阵的逆仍是初等矩阵的性质，我们得到：

$$(E_k E_{k-1} \dots E_2 E_1)^{-1} = (A^{-1})^{-1}$$

$$E_1^{-1} E_2^{-1} \dots E_{k-1}^{-1} E_k^{-1} = A$$

由于每个 E_i^{-1} 仍然是同类型的初等矩阵，上式表明：

任何可逆矩阵 A 都可以分解为一系列初等矩阵的乘积。

4.9 LU/LDU Decomposition

在高斯消元法中，我们通过一系列行变换将系数矩阵 A 化为上三角矩阵 U。这个过程可以视为用一系列初等矩阵左乘 A。而之前我们学过的一个重要定理是：

任何可逆矩阵A都可以分解为一系列初等矩阵的乘积。

这为我们提供了从矩阵乘法视角理解消元过程的理论基础。那么，这种初等矩阵的分解能否以一种更结构化、更实用的形式呈现呢？答案是肯定的——这就是本节要介绍的 LU **分解**。

考虑矩阵

$$A = \begin{bmatrix} 2 & 1 \\ 6 & 8 \end{bmatrix}.$$

我们的目标是将 6 消为 0。这一步消元操作可以用一个初等矩阵 E_{21} 表示：

$$E_{21} = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}, \quad \text{其中乘数}\ell_{21} = 3.$$

左乘 E_{21} 即完成消元：

$$E_{21}A = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 6 & 8 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 0 & 5 \end{bmatrix} = U.$$

现在，我们问：如何从 U 变回 A？这需要左乘 E_{21} 的逆矩阵：

$$E_{21}^{-1} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}.$$

容易验证：

$$E_{21}^{-1}U = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 6 & 8 \end{bmatrix} = A.$$

将 E_{21}^{-1} 记作 L（Lower triangular），我们就得到了

$$A = LU.$$

这就是 A 的 LU **分解**：一个**下三角矩阵** L（对角元为 1）乘以一个**上三角矩阵** U（对角元为消元后的主元）。

对于 $n \times n$ 矩阵 A（假设无需行交换），消元过程是一系列初等行变换 $E_{21}, E_{31}, \dots, E_{n,n-1}$ 的叠加。最终结果是：

$$(E_{n,n-1} \cdots E_{31} E_{21}) A = U.$$

将初等矩阵移到等式右边（注意顺序反转）：

$$A = (E_{21}^{-1} E_{31}^{-1} \cdots E_{n,n-1}^{-1}) U.$$

这正是"可逆矩阵可分解为初等矩阵乘积"定理在**无行交换情况下的具体实现**！定义

$$L = E_{21}^{-1} E_{31}^{-1} \cdots E_{n,n-1}^{-1},$$

我们便得到一般的分解式：

$$\boxed{A = LU}.$$

其中：

- L 是**单位下三角矩阵**（主对角线全为 1，下三角部分存储消元所用的**乘数** ℓ_{ij}）。
- U 是**上三角矩阵**（主对角线为消元后的主元，即高斯消元的最终结果）。

分解类型	**存在条件**	**与初等矩阵分解的关系**
初等矩阵分解	**A 必须可逆**	$A = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$（一般形式）
LU 分解（无行交换）	**A 顺序主子式非零**	特殊情形：所有 E_i^{-1} 恰好可合并为一个下三角矩阵 L
PA = LU 分解	**A 可为任意方阵**	一般情形：加入排列矩阵 P（也是初等矩阵乘积）

表 1: 不同分解的关系与条件

关键区别与联系：

1. **初等矩阵分解定理**是更基本的理论结果，它告诉我们可逆矩阵总可以写成初等矩阵的乘积，但这是一种"松散"的分解形式。
2. **LU 分解（无行交换）**是初等矩阵分解的一种**高度结构化特例**：当消元过程无需行交换，且所有顺序主子式非零时，那些初等逆矩阵 E_{ij}^{-1} 的乘积可以"打包"成一个下三角矩阵 L。
3. **顺序主子式条件**比可逆性更强：即使 A 可逆，也可能需要行交换才能进行 LU 分解。例如 $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ 是可逆的，但第一步主元为 0，需要行交换。
4. **$PA = LU$ 分解**是最实用的形式：通过引入排列矩阵 P（记录行交换），任何可逆矩阵都有 $PA = LU$ 分解。对于奇异矩阵，也能得到 $PA = LU$，但 U 会有零行。

例 1：对矩阵

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

进行消元：

1. 第 1 步：用乘数 $\ell_{21} = \frac{1}{2}$ 消去 (2,1) 元素。
2. 第 2 步：用乘数 $\ell_{32} = \frac{2}{3}$ 消去 (3,2) 元素。

得到分解：

$$A = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & \frac{2}{3} & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} 2 & 1 & 0 \\ 0 & \frac{3}{2} & 1 \\ 0 & 0 & \frac{4}{3} \end{bmatrix}}_{U}.$$

注意 $A(3,1) = 0$，所以不需要消元，对应 $L(3,1) = 0$。

例 2（奇异矩阵的情况）：考虑不可逆矩阵

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}.$$

消元：第二行减去 2 倍第一行：

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} = LU.$$

这里 U 有零行，反映了 A 的奇异性。注意这个 A 虽然不可逆，但仍然有无行交换的 LU 分解（因为它满足顺序主子式非零：$1 \neq 0$）。

标准的 LU 分解形式上"不对称"：U 的对角线是主元，而 L 的对角线是 1。我们可以通过引入一个对角矩阵 D 来平衡这种不对称性，得到更对称的 LDU **分解**。

将 U 分解为 D 和新的 U 的乘积：

$$U = D \cdot U_{\text{new}},$$

其中 D 包含所有主元 $d_1, d_2, \ldots, d_n$，而新的 U_{new} 对角线变为 1（即将 U 的每一行除以其主元）。于是：

$$A = L(DU_{\text{new}}) = (LD)U_{\text{new}}.$$

但更常见的对称写法是保留 L 为单位下三角，而将 D 置于中间：

$$\boxed{A = LDU}.$$

在 LDU 分解中，L 和 U 都是单位三角矩阵（对角元全为 1），D 为对角矩阵。

例子：回顾最初的 2×2 分解：

$$\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 8 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}.$$

这里 $D = \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}$，新的 $U = \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}$。

4.10 Rank of Matrix

矩阵的秩是线性代数中的一个核心概念，它深刻地反映了矩阵的本质特性。对于一个 $m \times n$ 矩阵 A，其**秩**（Rank）记作 $\rho(A)$ 或 $\text{rank}(A)$，可以通过多种等价的方式定义，这些定义从不同角度揭示了秩的几何与代数内涵：

- 秩是矩阵对应的线性方程组中**有效方程**（独立方程）的个数。
- 秩是矩阵化为**行简化阶梯形**（Reduced Row Echelon Form, RREF）后，其**首项 1**（leading 1）的个数。
- 秩是矩阵中非零**子式**（行列式）的最高阶数。
- 秩是矩阵的列向量组（或行向量组）中**线性无关向量**的最大个数，即列空间（或行空间）的维数。

矩阵的秩具有以下基本性质：

1. **有界性**：对于任意 $m \times n$ 矩阵 A，其秩满足：

$$0 \leq \rho(A) \leq \min\{m, n\}.$$

2. **转置不变性**：矩阵的秩在转置操作下保持不变，并且与相关乘积矩阵的秩相等：

$$\rho(A^T) = \rho(A) = \rho(A^T A) = \rho(AA^T).$$

3. **初等变换不变性**：矩阵的秩在初等变换下保持不变。若矩阵 A 与 B **等价**（即存在可逆矩阵 P, Q 使得 $B = PAQ$），记作 $A \sim B$，则有：

$$\rho(A) = \rho(B).$$

特别地，若 P, Q 可逆，则 $\rho(PAQ) = \rho(A)$。初等变换是矩阵等价的核心操作。

矩阵的秩满足一系列重要的不等式关系，在处理复合矩阵时尤为有用：

1. **分块矩阵的秩**：对于将矩阵 A, B 水平拼接得到的分块矩阵 (A, B)，其秩满足：

$$\max\{\rho(A), \rho(B)\} \leq \rho(A, B) \leq \rho(A) + \rho(B).$$

2. **矩阵和的秩**：

$$\rho(A + B) \leq \rho(A) + \rho(B).$$

3. **矩阵乘积的秩**（秩越乘越小）：

$$\rho(AB) \leq \min\{\rho(A), \rho(B)\}.$$

4. **Sylvester 秩不等式**：若 $A_{m \times n} B_{n \times p} = 0$，则

$$\rho(A) + \rho(B) \leq n.$$

5. **Frobenius 秩不等式**：

$$\rho(AB) + \rho(BC) \leq \rho(ABC) + \rho(B).$$

这些不等式可以直观地总结为：**“秩越乘越小，越拼越大，分开加最大”**。其中，“拼”指矩阵的拼接（分块），“加”指矩阵的加法。

$$①\text{解的情况的判别}\begin{cases}\text{无解} \Leftrightarrow \rho(A) < \rho(A|\mathbf{b}) \\ \text{有解} \Leftrightarrow \rho(A) = \rho(A|\mathbf{b}) \begin{cases}\text{唯一解} \Leftrightarrow \rho(A) = \rho(A|\mathbf{b}) = n\text{（列满秩）} \\ \text{无穷解} \Leftrightarrow \rho(A) = \rho(A|\mathbf{b}) < n\end{cases}\end{cases}$$

* $\rho(A) = m$（行满秩）$\Leftrightarrow \rho(A) = \rho(A|\mathbf{b})\ \forall\, \mathbf{b} \Leftrightarrow$ 总是有解

（$\rho(A) < m \Leftrightarrow$ 对某些$\mathbf{b}$有解，对某些$\mathbf{b}$无解）

$$\text{如果 } A: n\times n\text{方阵}\begin{cases}\text{行满秩/列满秩/满秩} \Leftrightarrow \text{对所有}\mathbf{b}\text{都有唯一解 } \mathbf{x} = A^{-1}\mathbf{b} \Leftrightarrow A^{-1}\text{存在} \Leftrightarrow |A| \neq 0 \\ \text{不满秩} \Leftrightarrow \text{对某些}\mathbf{b}\text{无解，对某些}\mathbf{b}\text{无穷解}\end{cases}$$

$$②\text{求解}\begin{cases}\text{一般解法：高斯消元 把增广矩阵化为 RRE} \\ \text{特殊情况：}A\text{是方阵（}m=n\text{）}\begin{cases}\text{有唯一解}\begin{cases}\text{先求}A^{-1}\text{，再乘}\mathbf{b}\begin{cases}\text{先求}|A|\text{和}adjA,\ A^{-1} = \dfrac{1}{|A|}adjA \\ \text{用初等行变换求 } A^{-1}: (A|I_n) \to (I_n|A^{-1})\end{cases} \\ \text{用 Cramer' s rule 分别求 } x_i\end{cases} \\ \text{无唯一解：用一般解法}\end{cases}\end{cases}$$

图 6: 线性系统解的情况判别以及求解总结

4.11 Exercises

5 Vector Spaces 向量空间

5.1 Vector Space and Subspace

本章让我们暂时忘掉矩阵和线性系统，回到向量的世界。之前我们介绍了什么是向量，并且在欧几里得空间 $\mathbb{R}^2$ 和 $\mathbb{R}^3$ 里对向量做了一些可视化的讨论。现在我们重新规范地定义一下向量存在的场域——**向量空间**（又叫**线性空间**）。$\mathbb{R}^2$ 和 $\mathbb{R}^3$ 都是向量空间的例子。

一个**实向量空间**（real vector space）V 是一个非空集合，配备有向量加法运算和标量乘法运算，使得对于所有 $\alpha, \beta \in \mathbb{R}$ 和所有 $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$，以下公理成立：

1. $\mathbf{u} + \mathbf{v} \in V$ (**加法封闭性**)
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (**加法交换律**)
3. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ (**加法结合律**)
4. 存在唯一的 $\mathbf{0} \in V$，称为**零向量**，使得对所有 $\mathbf{v} \in V$，有 $\mathbf{v} + \mathbf{0} = \mathbf{v}$
5. 对每个 $\mathbf{v} \in V$，存在一个元素 $\mathbf{w} \in V$（通常记为 $-\mathbf{v}$，称为 $\mathbf{v}$ 的**负元**），使得 $\mathbf{v} + \mathbf{w} = \mathbf{0}$
6. $\alpha\mathbf{v} \in V$ (**标量乘法封闭性**)
7. $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$ (**分配律**)
8. $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$ (**分配律**)
9. $\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}$ (**标量乘法结合律**)
10. $1\mathbf{v} = \mathbf{v}$

设 V 是一个向量空间。那么 V 的一个非空子集 W 是 V 的一个**子空间**（subspace），当且仅当以下两个条件同时成立：

- 对所有 $\mathbf{u}, \mathbf{v} \in W$，有 $\mathbf{u} + \mathbf{v} \in W$（即 W 对向量加法封闭）
- 对所有 $\mathbf{v} \in W$ 和所有 $\alpha \in \mathbb{R}$，有 $\alpha\mathbf{v} \in W$（即 W 对标量乘法封闭）

上述定理称为**子空间判别准则**（subspace criterion）。它提供了一个实用的方法来判定一个向量空间的子集本身是否构成一个向量空间。

5.2 Linear Span and Linear Independence

线性张成（linear span）集合 $X = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ 的线性张成，记为 $\mathrm{Lin}(X)$ 或 $\mathrm{Lin}\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$，是指向量 $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k$ 的所有线性组合构成的集合。即：

$$\mathrm{Lin}\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\} = \{\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k \mid \alpha_1, \alpha_2, ..., \alpha_k \in \mathbb{R}\}$$

设 V 是一个向量空间，且 $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k \in V$。向量 $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k$ 被称为**线性无关**（linearly independent）当且仅当向量方程

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k = \mathbf{0}$$

有唯一解 $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$。

现在有 m 个 $\mathbb{R}^n$ 里的向量，想要生成 $\mathbb{R}^n$：

- 如果 $m < n$，则这 m 个向量必不可能张成 $\mathbb{R}^n$，不确定是否线性无关
- 如果 $m = n$，且这 m 个向量线性无关，则一定可以张成 $\mathbb{R}^n$
- 如果 $m = n$，且这 m 个向量线性相关，则一定不可以张成 $\mathbb{R}^n$
- 如果 $m > n$，则这 m 个向量一定线性相关，但不确定是否张成 $\mathbb{R}^n$

向量构成的矩阵的秩就是有效向量的个数。

怎么判断一个向量组是否能张成 $\mathbb{R}^n$ 以及是否线性无关？我们可以用之前学习的线性系统的知识解答这个问题：

- 一个向量组能否张成 $\mathbb{R}^n$ 等价于：对于任意的 $\mathbf{b} \in \mathbb{R}^n$，方程组 $A\mathbf{x} = \mathbf{b}$ 都有解。其充要条件是 A 是行满秩的，即该向量组组成的矩阵是行满秩的。那么我们只需要对它做一次 RRE 即可判断。
- 线性无关按照定义等价于齐次线性系统 $A\mathbf{x} = \mathbf{0}$ 有唯一解。而我们知道齐次线性系统一定有解，而已知有解的情况下，唯一解的充要条件是列满秩，因此就是检查这个由向量组组成的矩阵是否是列满秩的。

若向量组内存在一个部分组，满足 (1) 该部分组线性无关；(2) 原向量组中的任一向量都能由该部分组线性表示，则称该部分组是原向量组的一个**极大线性无关组**（maximal linearly independent group）。

初等行变换不会改变列向量之间的线性关系。

5.3 Basis and Dimension

设 V 是一个向量空间。V 的一个子集 $B = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ 称为 V 的一个**基**（basis），当且仅当 B 是一个线性无关集并且能张成 V。

一个向量空间的基中所含向量的个数，称为这个空间的**维数**（dimension），记作 $\dim(V)$。生成这个空间所需的最少向量个数，就是该空间的维数。

$$\dim(\mathbb{R}^n) = n$$

5.4 Normed Vector Space and Inner Product Space

在欧几里得空间 $\mathbb{R}^n$ 里，我们定义了向量的点积（dot product），继而定义了向量的长度（norm）。我们现在知道了向量空间未必是 $\mathbb{R}^n$，可以是很抽象的空间，因此我们现在把点积的概念一般化，拓展到任意向量空间中的**内积**（inner product）概念，用符号 $\langle \mathbf{a}, \mathbf{b} \rangle$ 表示。

内积可以任意定义，只需要满足以下三个条件：对于任意向量 $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ 和任意标量 $\alpha \in \mathbb{R}$：

1. **对称性**（Symmetry）：$\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. **线性性**（Linearity）：$\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w} \rangle = \alpha\langle \mathbf{u}, \mathbf{w} \rangle + \beta\langle \mathbf{v}, \mathbf{w} \rangle$
3. **正定性**（Positive Definiteness）：
   - $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$
   - $\langle \mathbf{u}, \mathbf{u} \rangle = 0 \iff \mathbf{u} = \mathbf{0}$

上述条件也可以把“线性性”改为“双线性性”，但注意对称和线性可以推出双线性。

装备了内积的向量空间就被称为**内积空间**（inner product space），此时该向量空间才有了距离、长度、角度的概念。内积是一个比单纯的范数更强的定义，有了内积我们就可以自然地定义范数是一个向量与自身内积的算术平方根，因此一个内积空间一定是一个**赋范向量空间**（normed vector space）。而如果一个向量空间只被装备了范数定义而无内积定义，则它是一个赋范向量空间但不是一个内积空间，它仍然可以讨论向量的长度和两个点之间的距离，但没有角度的概念。另外，可以看出在向量空间中两个点之间的距离的定义又弱于向量的范数的定义，因为向量空间中两个点之间的距离可以看作一个向量的范数，因此定义了内积就自然定义了范数，而定义了范数就自然定义了距离。当然，反过来说，距离是一个更底层更基础的概念，定义在向量空间中的范数是距离定义的一种特例。定义了距离的**拓扑空间**（topological space）被称为**度量空间**（metric space）。

我们已经知道，欧几里得空间 $\mathbb{R}^n$ 本身就是一个内积空间，它自带的标准内积是**点积**（dot product），也称为标量积（scalar product）。对于 $\mathbb{R}^n$ 中的向量 $\mathbf{a} = (a_1, a_2, \ldots, a_n)^T$ 和 $\mathbf{b} = (b_1, b_2, \ldots, b_n)^T$：

$$\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T\mathbf{b} = \sum_{i=1}^{n} a_i b_i$$

通过内积，我们可以定义向量的**长度**（范数）：

$$\|\mathbf{a}\| = \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle}$$

以及两个向量之间的**距离**：

$$d(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\|$$

我们已经说过了**内积**的几何意义是一个向量在另一个向量上的正交投影的长度乘以那个向量的长度。注意如果那另一个向量是单位向量，那我们会得到更加便捷的几何意义，即内积就直接等于投影的长度。内积的“投影”意义非常重要，在之后我们会多次用到。

在内积空间中，我们可以定义**正交**的概念：

- 两个向量 $\mathbf{u}, \mathbf{v}$ 是**正交**的（orthogonal），如果 $\langle \mathbf{u}, \mathbf{v} \rangle = 0$
- 一个向量集合是**正交向量组**，如果其中任意两个不同向量都正交
- 一个向量集合是**标准正交组**（orthonormal），如果它是正交向量组且每个向量的范数都是 1

Gram-Schmidt 正交化过程是将一组线性无关的向量转化为标准正交组的方法。给定线性无关的向量组 $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$，我们可以通过以下步骤构造标准正交组 $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$：

$$\begin{aligned}
\mathbf{u}_1 &= \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} \\
\mathbf{u}_2 &= \frac{\mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{u}_1 \rangle \mathbf{u}_1}{\|\mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{u}_1 \rangle \mathbf{u}_1\|} \\
\mathbf{u}_3 &= \frac{\mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{u}_1 \rangle \mathbf{u}_1 - \langle \mathbf{v}_3, \mathbf{u}_2 \rangle \mathbf{u}_2}{\|\mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{u}_1 \rangle \mathbf{u}_1 - \langle \mathbf{v}_3, \mathbf{u}_2 \rangle \mathbf{u}_2\|} \\
&\vdots
\end{aligned}$$

一般地，对于 $i = 1, 2, \ldots, k$：

$$\mathbf{u}_i = \frac{\mathbf{v}_i - \sum_{j=1}^{i-1} \langle \mathbf{v}_i, \mathbf{u}_j \rangle \mathbf{u}_j}{\left\|\mathbf{v}_i - \sum_{j=1}^{i-1} \langle \mathbf{v}_i, \mathbf{u}_j \rangle \mathbf{u}_j\right\|}$$

这个过程保持了线性张成关系：对于每个 i，有 $\mathrm{Lin}\{\mathbf{u}_1, \ldots, \mathbf{u}_i\} = \mathrm{Lin}\{\mathbf{v}_1, \ldots, \mathbf{v}_i\}$。

5.5 Hilbert Space

如果一个**内积空间**还具有**完备性**（completeness），则被称为**希尔伯特空间**（Hilbert space）。希尔伯特空间允许无限维，尽管本笔记不会讨论无限维空间的性质（参考泛函分析）。

- 注意，希尔伯特空间并非必须是无限维，有限维的希尔伯特空间是完全合法的，因为 "一个完备的内积空间" 是它的全部定义。
- 复习一下完备性的含义：这个空间的任何柯西序列都必须在该空间内收敛到一个极限。注意完备性是一个度量/拓扑性质，与维度无关。
- 在有限维，任何定义了范数（由内积诱导）的有限维内积空间自动是完备的。所以，$\mathbb{R}^3$ 或 $\mathbb{C}^{100}$ 不仅是内积空间，它们天然就是希尔伯特空间。在无限维，完备性就不是自动满足的了。

我们知道实数系是完备的，因此装备了标准点积的欧几里得空间 $\mathbb{R}^n$ 不仅是一个内积空间还是一个希尔伯特空间。注意，它是一个**有限维希尔伯特空间**。

希尔伯特空间是一类非常重要的空间，我们之后会频繁地讨论它。除了我们介绍的这些空间，数学中的抽象空间类型还有很多，例如**巴拿赫空间**（Banach space）、**局部凸空间**（locally convex space）等等。下图总结了一些常见的抽象空间类型的从属关系，以及本笔记的重点研究对象——欧几里得空间 $\mathbb{R}^n$ 在其中的位置。（A 箭头指向 B 表示一个 A 空间也是一个 B 空间。）

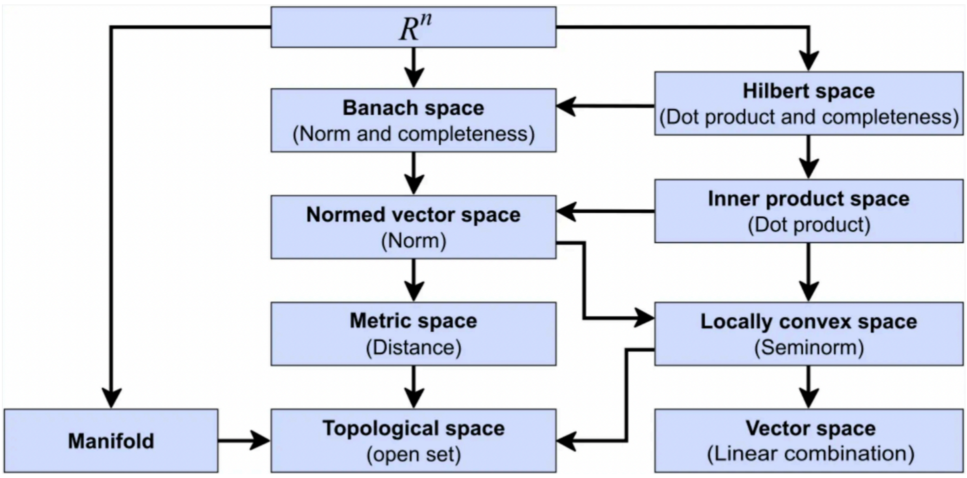

图 7: 抽象空间类型概述

5.6 Coordinates with respect to a basis

$\mathbb{R}^n$ 的**标准基 (standard basis)** 记为 $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$。该基中的向量定义为：

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \cdots \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

对于任意向量 $\mathbf{v} \in \mathbb{R}^n$，其中 $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$，我们知道它可以唯一地表示为基向量的线性组合。引

入标量 $\alpha_1,\alpha_2,\ldots,\alpha_n$，有：

$$\mathbf{v}=\alpha_1\mathbf{e}_1+\alpha_2\mathbf{e}_2+\cdots+\alpha_n\mathbf{e}_n,$$

或者等价地，

$$\begin{pmatrix}v_1\\v_2\\\vdots\\v_n\end{pmatrix}=\alpha_1\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}+\alpha_2\begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}+\cdots+\alpha_n\begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}.$$

由此可得 $\alpha_1=v_1,\alpha_2=v_2,\ldots,\alpha_n=v_n$。标量 $\alpha_1,\alpha_2,\ldots,\alpha_n$ 被称为向量 $\mathbf{v}$ 关于标准基的**标准坐标 (standard coordinates)**，记为 $(\mathbf{v})$。当然，向量 $\mathbf{v}$ 关于标准基的坐标 $(\mathbf{v})$ 在数值上与其作为 $\mathbb{R}^n$ 中向量的分量是一致的；即：

$$(\mathbf{v})=\begin{pmatrix}v_1\\v_2\\\vdots\\v_n\end{pmatrix}=\mathbf{v}.$$

然而，对于 $\mathbb{R}^n$ 的其他基，情况并非如此。

一般来说，如果 $B=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_k\}$ 是一个 k 维向量空间 V 的一组基，且 $\mathbf{v}$ 是 V 中的任意向量，那么在唯一的线性组合

$$\mathbf{v}=\alpha_1\mathbf{v}_1+\alpha_2\mathbf{v}_2+\cdots+\alpha_k\mathbf{v}_k$$

中出现的标量 $\alpha_1,\alpha_2,\ldots,\alpha_k$，被称为向量 $\mathbf{v}$ 关于基 B 的**坐标 (coordinates)**，并记为 $(\mathbf{v})_B$。如果基 B 恰好是标准基，下标 B 可以省略。

注意区分标准基（standard basis），标准正交基（orthonormal basis）和正交基（orthogonal basis)。虽然“标准”在“标准正交基”中代表“单位长度”的意思，但想表达都是单位长度但未必正交的一组基却不能直接说“标准基”，因为另一个概念占用了这个中文译名，可以直接说“单位基/单位长度基（unit basis)”或者“归一化基（normalized basis)”。它们的关系如下：标准基 $\subset$ 标准正交基 $\subset$ 正交基 $\subset$ 基。

在已知空间的一组基 $B=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$ 后，如何求向量 $\mathbf{x}$ 在此基下的坐标 $[\alpha_1,\alpha_2,\ldots,\alpha_n]^T$，即求解 $\mathbf{x}=\alpha_1\mathbf{v}_1+\alpha_2\mathbf{v}_2+\cdots+\alpha_n\mathbf{v}_n$，其方法因基的性质而异。

对于任意一组基（线性无关即可），最普适的方法是求解线性方程组。

- 将坐标未知的方程 $\mathbf{x}=\sum_{i=1}^n\alpha_i\mathbf{v}_i$ 具体写开。
- 将其转化为关于 $\alpha_1,\alpha_2,\ldots,\alpha_n$ 的线性方程组。
- 由于 $\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$ 是基，该方程组有唯一解，可通过高斯消元法求解。

此方法万无一失，但计算量可能较大。

例题 1：一般基（需解方程组）在 $\mathbb{R}^2$ 中，给定标准基 $B=\{\mathbf{e}_1,\mathbf{e}_2\}$ 和非标准基 $C=\{\mathbf{v}_1,\mathbf{v}_2\}$，其中：

$$\mathbf{e}_1=\begin{pmatrix}1\\0\end{pmatrix},\quad\mathbf{e}_2=\begin{pmatrix}0\\1\end{pmatrix},\quad\mathbf{v}_1=\begin{pmatrix}1\\2\end{pmatrix},\quad\mathbf{v}_2=\begin{pmatrix}1\\-1\end{pmatrix}.$$

向量 $\mathbf{v}=\begin{pmatrix}2\\-5\end{pmatrix}$ 在标准基 B 下的坐标为：

$$(\mathbf{v})=\begin{pmatrix}2\\-5\end{pmatrix}.$$

现在求 $\mathbf{v}$ 在基 C 下的坐标 $(\mathbf{v})_C = \begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}_C$，即解方程：

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 = \mathbf{v}$$

$$\alpha_1\begin{pmatrix}1\\2\end{pmatrix} + \alpha_2\begin{pmatrix}1\\-1\end{pmatrix} = \begin{pmatrix}2\\-5\end{pmatrix}$$

这等价于求解线性方程组：

$$\begin{pmatrix}1 & 1\\ 2 & -1\end{pmatrix}\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix} = \begin{pmatrix}2\\-5\end{pmatrix}$$

通过高斯消元法求解，得到 $\alpha_1 = -1, \alpha_2 = 3$。因此：

$$(\mathbf{v})_C = \begin{pmatrix}-1\\3\end{pmatrix}_C$$

验证：$-1\cdot\begin{pmatrix}1\\2\end{pmatrix} + 3\cdot\begin{pmatrix}1\\-1\end{pmatrix} = \begin{pmatrix}-1+3\\-2-3\end{pmatrix} = \begin{pmatrix}2\\-5\end{pmatrix}$。

若基 B 是**正交基 (Orthogonal Basis)**，即满足 $i \neq j$ 时 $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = 0$，则计算将大大简化。

为了求坐标 α_k，我们将等式 $\mathbf{x} = \sum_{i=1}^{n}\alpha_i\mathbf{v}_i$ 的两端与基向量 $\mathbf{v}_k$ 做内积：

$$\langle \mathbf{x}, \mathbf{v}_k\rangle = \left\langle \sum_{i=1}^{n}\alpha_i\mathbf{v}_i, \mathbf{v}_k\right\rangle$$

利用内积的线性性质及正交性（当 $i \neq k$ 时 $\langle \mathbf{v}_i, \mathbf{v}_k\rangle = 0$），上式右边只剩下第 k 项：

$$\langle \mathbf{x}, \mathbf{v}_k\rangle = \alpha_k\langle \mathbf{v}_k, \mathbf{v}_k\rangle = \alpha_k\|\mathbf{v}_k\|^2$$

由此立即可得：

$$\alpha_k = \frac{\langle \mathbf{x}, \mathbf{v}_k\rangle}{\langle \mathbf{v}_k, \mathbf{v}_k\rangle} = \frac{\langle \mathbf{x}, \mathbf{v}_k\rangle}{\|\mathbf{v}_k\|^2}$$

上述公式有清晰的几何意义：

- 分子 $\langle \mathbf{x}, \mathbf{v}_k\rangle$ 是向量 $\mathbf{x}$ 在基向量 $\mathbf{v}_k$ 方向上的**投影长度**与 $\mathbf{v}_k$ 本身长度的乘积。
- 比值 $\frac{\langle \mathbf{x}, \mathbf{v}_k\rangle}{\|\mathbf{v}_k\|}$ 得到的是 $\mathbf{x}$ 在 $\mathbf{v}_k$ 方向上的**投影长度**。
- 由于坐标 α_k 的定义就是"需要多少个 $\mathbf{v}_k$ 单位向量才能拼出投影分量"，因此需要将投影长度再除以坐标轴一个单位的长度（即 $\|\mathbf{v}_k\|$）。整个公式 $\frac{\langle \mathbf{x}, \mathbf{v}_k\rangle}{\|\mathbf{v}_k\|^2}$ 等价于 $\frac{\text{投影长度}}{\|\mathbf{v}_k\|}$，完美地给出了坐标值。

例题 2：正交基（使用内积公式） 考虑 $\mathbb{R}^3$ 中的向量 $\mathbf{v} = \begin{pmatrix}2\\4\\6\end{pmatrix}$ 和正交基 $B = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$，其中：

$$\mathbf{u}_1 = \begin{pmatrix}2\\0\\0\end{pmatrix},\quad \mathbf{u}_2 = \begin{pmatrix}0\\2\\0\end{pmatrix},\quad \mathbf{u}_3 = \begin{pmatrix}0\\0\\2\end{pmatrix}.$$

容易验证 $\langle \mathbf{u}_i, \mathbf{u}_j\rangle = 0$（当 $i \neq j$ 时），且 $\|\mathbf{u}_1\| = \|\mathbf{u}_2\| = \|\mathbf{u}_3\| = 2$。使用正交基下的坐标公式 $\alpha_k = \frac{\langle \mathbf{v}, \mathbf{u}_k\rangle}{\|\mathbf{u}_k\|^2}$ 进行计算：

$$\alpha_1 = \frac{\langle \mathbf{v}, \mathbf{u}_1\rangle}{\|\mathbf{u}_1\|^2} = \frac{2\times 2 + 4\times 0 + 6\times 0}{2^2} = \frac{4}{4} = 1$$

$$\alpha_2 = \frac{\langle \mathbf{v}, \mathbf{u}_2 \rangle}{\|\mathbf{u}_2\|^2} = \frac{2 \times 0 + 4 \times 2 + 6 \times 0}{2^2} = \frac{8}{4} = 2$$
$$\alpha_3 = \frac{\langle \mathbf{v}, \mathbf{u}_3 \rangle}{\|\mathbf{u}_3\|^2} = \frac{2 \times 0 + 4 \times 0 + 6 \times 2}{2^2} = \frac{12}{4} = 3$$

因此，向量 $\mathbf{v}$ 在正交基 B 下的坐标为：

$$(\mathbf{v})_B = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}_B$$

验证：$1 \cdot \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} + 2 \cdot \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} + 3 \cdot \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}$。

若基 B 是**标准正交基 (Orthonormal Basis)**，即满足 $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \delta_{ij}$，则公式简化为最优雅的形式：

$$\alpha_k = \langle \mathbf{x}, \mathbf{v}_k \rangle$$

这意味着在标准正交基下，向量的坐标直接等于它与各基向量的内积。这是正交基情形的特例，因为此时 $\|\mathbf{v}_k\| = 1$，分母为 1。

例题 3：标准正交基（坐标即内积） 考虑 $\mathbb{R}^2$ 中的向量 $\mathbf{w} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ 和标准正交基 $B = \{\mathbf{q}_1, \mathbf{q}_2\}$，其中：

$$\mathbf{q}_1 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}, \quad \mathbf{q}_2 = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}.$$

容易验证 $\|\mathbf{q}_1\| = \|\mathbf{q}_2\| = 1$ 且 $\langle \mathbf{q}_1, \mathbf{q}_2 \rangle = 0$。在标准正交基下，坐标直接等于向量与基向量的内积：

$$\beta_1 = \langle \mathbf{w}, \mathbf{q}_1 \rangle = 3 \times \frac{\sqrt{2}}{2} + 4 \times \frac{\sqrt{2}}{2} = \frac{7\sqrt{2}}{2}$$
$$\beta_2 = \langle \mathbf{w}, \mathbf{q}_2 \rangle = 3 \times \left(-\frac{\sqrt{2}}{2}\right) + 4 \times \frac{\sqrt{2}}{2} = \frac{\sqrt{2}}{2}$$

因此，向量 $\mathbf{w}$ 在标准正交基 B 下的坐标为：

$$(\mathbf{w})_B = \begin{pmatrix} \frac{7\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}_B$$

验证：$\frac{7\sqrt{2}}{2} \cdot \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} + \frac{\sqrt{2}}{2} \cdot \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} = \begin{pmatrix} \frac{7}{2} - \frac{1}{2} \\ \frac{7}{2} + \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$。

基的类型	**推荐方法**	**计算复杂度**
一般基	解线性方程组	$O(n^3)$ (如高斯消元)
正交基	使用公式 $\alpha_k = \frac{\langle \mathbf{x}, \mathbf{v}_k \rangle}{\|\mathbf{v}_k\|^2}$	$O(n^2)$ (n 个内积)
标准正交基	使用公式 $\alpha_k = \langle \mathbf{x}, \mathbf{v}_k \rangle$	$O(n^2)$ (n 个内积)

因此，在处理坐标问题时，如果已知基是正交的或标准正交的，务必利用内积公式来简化计算，这不仅是计算技巧，也体现了这些特殊基的优良性质。

无论选择哪种基，只要它是一组基，就必然满足两个核心性质：

- **完备性 (Completeness)**：空间中的**每一个**向量都一定能用该基的线性组合来表示。
- **坐标唯一性 (Uniqueness of Coordinates)**：对于空间中的**每一个**向量，其在这种基下的坐标表示是**唯一**的。

5.7 Orthogonality between Subspaces

两个子空间 S 和 T 正交意味着 S 中任意向量和 T 中任意向量都正交。
例如，在 $\mathbb{R}^3$ 中：

- z 轴和 x 轴是两个正交的子空间
- z 轴和 x-y 平面也是正交的子空间

在 $\mathbb{R}^4$ 中，两个不同的平面也可能正交，比如 $\mathrm{Lin}\{(1,0,0,0)\}$ 和 $\mathrm{Lin}\{(0,1,0,0)\}$。

- 两个正交子空间的交点只能是零向量。（例如，地板和墙壁不正交，因为它们相交于一条直线）
- 两个正交子空间的基合在一起依然线性无关。

注意，当我们说“两个子空间互相正交”时，这两个子空间必须属于同一个母空间（比如都在 $\mathbb{R}^n$ 中），否则“正交”这个概念就没有意义或不成立，因为你没法对它们之间的向量做内积运算。“正交”的定义依赖于向量之间的内积关系，而内积只有在相同空间中有意义。如果两个子空间分别属于不同的空间，即使它们的维度相同，内积结构一样（比如都是 Euclidean dot product），我们仍然不能说它们是否正交。考虑 $U \subseteq \mathbb{R}^m$, $V \subseteq \mathbb{R}^n$, $\dim U = \dim V = k$，这时候你是没法拿一个 $\mathbf{u} \in U$，一个 $\mathbf{v} \in V$，直接计算 $\mathbf{u} \cdot \mathbf{v} =?$ 因为它们维度都不一样，无法点乘。什么情况下可以比较两个子空间的正交性呢？你必须把它们嵌入到一个共同的空间中。比如：把 $\mathbb{R}^m$ 和 $\mathbb{R}^n$ 都嵌入到 $\mathbb{R}^{\max(m,n)}$，或者更正式地，把 U 和 V 映射到同一个内积空间 $W \subseteq \mathbb{R}^N$，然后再比较它们在这个共同空间中的正交性。

总结，两个子空间只有在属于同一个内积空间时，才能讨论它们是否正交。即使两个子空间维度相同，结构相似，如果它们属于不同的母空间，就无法定义向量之间的点积，自然也无法比较正交性。

如果子空间 A 和子空间 B 正交，这两个空间的维度有要求吗？还是任意维度都可能成立正交关系呢？答案是有要求。两个子空间要正交，它们的维度不能太高，具体来说，必须满足：

$$\dim(A) + \dim(B) \leq \dim(\text{母空间})$$

由这个结论也可以直接判断地板和墙壁不是 $\mathbb{R}^3$ 中正交的子空间。因为两个平面的维度和等于 4，大于母空间的维度 3，所以在 $\mathbb{R}^3$ 中，不可能存在两个彼此正交的平面。（严格证明：两个不同的平面必定相交于一条直线。这条交线中的非零向量同时属于 A 和 B，它不可能与自己正交。）

5.8 Sum and Intersection of Subspaces

设 U 和 W 是实向量空间 V 的两个子空间。定义它们的**和**为：

$$U + W = \{\mathbf{u} + \mathbf{w} \mid \mathbf{u} \in U \text{ 且} \mathbf{w} \in W\}$$

定义它们的**交**为：

$$U \cap W = \{\mathbf{x} \mid \mathbf{x} \in U \text{ 且} \mathbf{x} \in W\}$$

子空间的和空间是由各子空间的基向量合在一起所共同线性生成的空间。如果这些基向量是线性无关的，则两个子空间没有重叠部分，这等价于 $U \cap W = \{\mathbf{0}\}$，也等价于后面要讨论的直和。

若两个子空间公共的唯一向量为零向量，则称这两个子空间是**无交连**（disjoint）的。

设 W_1, W_2 为向量空间 V 的子空间，则以下三条件等价：

1. $W_1 \subseteq W_2$
2. $W_1 \cap W_2 = W_1$
3. $W_1 + W_2 = W_2$

设 W_1, W_2 为向量空间 V 的子空间，则成立以下重要的**维数公式**：

$$\dim W_1 + \dim W_2 = \dim(W_1 + W_2) + \dim(W_1 \cap W_2)$$

- 如果两个子空间的基线性无关，则维度的和等于和的维度（因为维度等于基向量的个数），此时交空间的维度为 0，即交空间只有零向量。这种情况就是所谓的“直和”。
- 如果两组基向量中存在线性相关的向量，则维度的和大于和的维度，多出的维度等于交空间的维度。
- 维度的和不可能小于和的维度。

和空间的关键在于基的合并与维度的计算，而不仅仅是两个生成空间的简单“相加”。

在子空间的和运算中，满足交换律和结合律：

- **交换律**：$W_1 + W_2 = W_2 + W_1$
- **结合律**：$(W_1 + W_2) + W_3 = W_1 + (W_2 + W_3)$

设 U 和 W 是实向量空间 V 的两个子空间。如果和 $U + W$ 中的每个向量都能**唯一地**表示为 U 中的一个向量与 W 中的一个向量之和，则称这个和为**直和**，记作 $U \oplus W$。

和 $U + W$ 是直和当且仅当 $U \cap W = \{\mathbf{0}\}$。

类型	表示	几何意义	交空间性质
非直和	$U + W$	生成空间有重合部分	$\dim(U \cap W) > 0$，交空间含非零向量
直和	$U \oplus W$	生成空间无重合部分	$U \cap W = \{\mathbf{0}\}$，交空间只有零向量

对于直和，显然有：

$$\dim(W_1 \oplus W_2) = \dim W_1 + \dim W_2$$

5.9 Complement and Orthogonal Complement of a Subspace

设 U 和 W 是实向量空间 V 的两个子空间。如果满足：

$$U \oplus W = V$$

则称 W 是 U 的**补空间**（complement）。

注意每一个子空间一定存在补空间，所以我们可以说，设 U 是 V 的一个子空间，那么存在子空间 W，使得 $U \oplus W = V$，称 W 为 U 的补空间。

V 的每个子空间 U 都有补空间。例如，如果 $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ 是 U 的一组基，我们可以将其扩展为 V 的一组基：

$$\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}_{k+1}, \mathbf{u}_{k+2}, ..., \mathbf{u}_{k+n}\}$$

那么子空间

$$W = \mathrm{Lin}\{\mathbf{u}_{k+1}, \mathbf{u}_{k+2}, ..., \mathbf{u}_{k+n}\}$$

就是 U 的一个补空间。

V 的每个非平凡真子空间 U 都可以有许多不同的补空间。

设 U 是实向量空间 V 的一个子空间。U 的**正交补**（orthogonal complement）定义为：

$$U^\perp = \{\mathbf{v} \in V \mid \langle \mathbf{u}, \mathbf{v} \rangle = 0 \text{ 对所有} \mathbf{u} \in U \text{ 成立}\}$$

如果 U 是有限维实向量空间 V 的一个子空间，那么：

1. $U^\perp$ 也是 V 的子空间
2. $V = U \oplus U^\perp$
3. $\dim(V) = \dim(U) + \dim(U^\perp)$
4. $(U^\perp)^\perp = U$

n 维欧几里得空间 V 的每一个子空间 V_1 都有**唯**一的正交补。

	一般补空间	**正交补空间**
存在性	存在	存在
唯一性	不唯一	唯一
构造方式	通过扩展基	通过正交条件
表示	$U \oplus W = V$	$U \oplus U^\perp = V$

5.10 Exercises

6 Vector Spaces in Matrices 矩阵中的向量空间

在第四章中，矩阵的出现是作为一个数据仓库来存储了方程组的所有数据信息，在这一章中，我们引出矩阵的第二个身份/作用：空间建筑师。我们说过矩阵可以看作一个向量组，那么有向量就会有向量生成的线性空间。

6.1 Column Space, Row Space, Null Space

矩阵可以看作是由行向量或者列向量构成的向量组，那么自然可以来研究它的生成空间，它可以为线性系统 $A\mathbf{x} = \mathbf{b}$ 提供一个几何视角的理解。

列空间 (column space): $CS(A) = \text{Lin}\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$

行空间 (row space): $RS(A) = \text{Lin}\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m\}$

零空间 (null space): $N(A) = \text{Lin}\{\mathbf{x} \mid A\mathbf{x} = \mathbf{0}\}$

$N(A)$ 的基：方程的解中的方向向量

$CS(A)$ 的基：矩阵中包含**主元 (leading one)** 的列

$RS(A)$ 的基：**简化行阶梯形 (RRE)** 矩阵中包含主元的行

回到线性系统，之前我们纯代数角度的理解是把矩阵 A 按行分开看的，一行是一个方程，主元的个数是有效方程的个数。现在我们尝试把矩阵 A 按列分开看，这等价于把线性方程组写成它的向量形式：

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} x_1 + \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} x_2 + \cdots + \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} x_n = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

你可以从线性方程组化到这一步，也可以从 $A\mathbf{x} = \mathbf{b}$ 矩阵的乘法理解角度化到这一步，都是等价的。

左边就成了矩阵 A 列向量的线性组合形式，左边可以取的值域就是当 x_i 可以取任意值时，也就是矩阵 A 列向量的所有线性组合，也就是它的列空间，所以 $CS(A) = \mathcal{R}(A)$。从这个空间的角度理解，系统有解的等价条件为向量 $\mathbf{b}$ 要在 $CS(A)$ 里，也就是往原本由矩阵 A 的列向量构成的向量组里新塞入向量 $\mathbf{b}$，向量组的生成空间不能拓展。此时秩可以看作矩阵的行或列向量的极大线性无关组的向量个数，也就是 CS 的维度，因此等价于 $\text{rank}(A) = \text{rank}(A|\mathbf{b})$。

$RS(A)$ 有什么用，注意，$\text{rank}(A) = \text{rank}(A^T)$，所以 RS 和 CS 中有效向量的个数是一样的，$\dim(RS(A)) = \dim(CS(A)) = \text{rank}(A)$。$RS$ 和 CS 两者维度都等于 $\text{rank}(A)$，但它们是不同空间中的子空间。CS 是位于 $\mathbb{R}^m$ 中的子空间，RS 是位于 $\mathbb{R}^n$ 中的子空间。除非 A 是方阵，否则它们根本不可能位于同一个向量空间。可以理解为：行空间和列空间是不同维度的向量空间中的“形状相同”的子空间。你可以想象，行空间是嵌在 $\mathbb{R}^n$ 里的一个 r-维平面。列空间是嵌在 $\mathbb{R}^m$ 里的一个 r-维平面。

从空间角度理解为什么行满秩等价于 $\text{rank}(A) = \text{rank}(A|\mathbf{b})$ for all $\mathbf{b}$

充分：行满秩意味着 $\text{rank}(A) = m$，意味着 A 的 CS 和 RS 都是整个 $\mathbb{R}^m$，而 $\mathbf{b}$ 向量来自 $\mathbb{R}^m$，因此 $\mathbf{b}$ 一定在 $CS(A)$ 中。

必要：假设 A 不是行满秩，即 $\text{rank}(A) < m$，说明 A 的行向量不能张成整个 $\mathbb{R}^m$。这意味着存在某个 $\mathbf{b}$ 不在 A 的行空间中。对于这个 $\mathbf{b}$，增广矩阵 $(A|\mathbf{b})$ 会有一个新的主元，使得 $\text{rank}(A|\mathbf{b}) > \text{rank}(A)$，从而违背了 $\text{rank}(A) = \text{rank}(A|\mathbf{b})$ 的假设。

下面讨论零空间，零空间是 $\mathbb{R}^n$ 的子空间。

$N(A)$ 是齐次系统 $A\mathbf{x}=\mathbf{0}$ 所有解的集合.

$\dim(RS(A))/\dim(CS(A))/\text{rank}(A)+\dim(N(A))/\text{nullity}(A)=n$

* 注意虽然 RS 和 CS 这两个子空间的维度是一样的，但 $\dim(CS(A))+\dim(N(A))=n$ 这个等式并没有什么几何上的意义，因为 CS 和 N 根本不是一个母空间中的子空间，这个式子相当于在把一个 $\mathbb{R}^m$ 中的子空间的维度与一个 $\mathbb{R}^n$ 中子空间的维度相加。

$\dim(N(A))$ 是自由参数的数量，也就是列向量组中非主元的个数。线性系统中解的唯一性取决于零空间的维度。如果 $\text{nullity}(A)=0$，则解唯一；否则，解有无穷多个。对于方阵 A，如果 $\text{rank}(A)=n$，则 $\text{nullity}(A)=0$，A 可逆。

这个定理揭示了矩阵的信息分布：$\text{rank}(A)$ 告诉我们矩阵 A 的信息量有多大，$\text{nullity}(A)$ 告诉我们有多少“冗余”在 A 中。

6.2 The Rank-Nullity Theorem

$$a+b=c$$

$$a:\dim(RS(A))=\dim(CS(A))=\text{rank}(A)$$

$$b:\dim(N(A))=\text{nullity}(A)$$

$$c:n=\text{number of columns of } A$$

6.3 The Structure of the Solutions of Linear Systems

假设 A 是一个 $m\times n$ 矩阵，且系统 $A\mathbf{x}=\mathbf{b}$ 是相容的。令 $\mathbf{p}$ 是该系统的一个解，即 $A\mathbf{p}=\mathbf{b}$。那么该系统的解集恰好由向量 $\mathbf{p}+\mathbf{z}$ 组成，其中 $\mathbf{z}\in N(A)$；即：

$$\{\mathbf{x}\mid A\mathbf{x}=\mathbf{b}\}=\{\mathbf{p}+\mathbf{z}\mid \mathbf{z}\in N(A)\}$$

这个结果被称为**线性性原理 (principle of linearity)**。它表明，相容线性系统 $A\mathbf{x}=\mathbf{b}$ 的通解 $\mathbf{x}$ 由该系统的一个特解 $\mathbf{p}$ 加上对应的齐次系统 $A\mathbf{z}=\mathbf{0}$ 的通解 $\mathbf{z}$ 给出；即 $\mathbf{x}=\mathbf{p}+\mathbf{z}$。

对于齐次线性方程组而言，它的解集就是 A 的零空间，是 $\mathbb{R}^n$ 中一个通过原点的线性子空间。对于非齐次线性方程组而言，如果有解，它的解集是一个平移了的零空间，我们称之为 $\mathbb{R}^n$ 中的一个**仿射子空间 (affine subspace)**，而不是线性子空间，因为它不一定过原点。更具体地说，它是 A 的零空间在 $\mathbb{R}^n$ 中沿某个特解方向的平移，所以它的几何结构和零空间一模一样，但位置不同。仿射子空间与线性子空间的区别在于：仿射子空间不一定经过原点，它不再对“数乘”封闭，但保留了“向量之间的方向结构”。线性系统的解集的方向结构由零空间决定，而位置由特解决定。

6.4 Orthogonality between $N(\mathbf{A})$ and $RS(\mathbf{A})$

$N(A)$ 中的任意向量与 $RS(A)$ 中的任意向量正交。$N(A)$ 和 $RS(A)$ 不仅正交，而且互为正交补，它们的直和等于 $\mathbb{R}^n$。

如果 A 是一个实的 $m\times n$ 矩阵，那么 $RS(A)^\perp=N(A)$。注意很多人初学时会搞混，记成 $N(A)$ 跟 $CS(A)$ 正交，但其实稍微想一想就能知道这个结论的荒谬之处：$N(A)$ 和 $CS(A)$ 根本不是同一个母空间的子空间，它们根本不在一个空间里，没有“正交”可言。

注意 $RS(A)$ 还可以写成 $CS(A^T)$，所以上面的等式还可以写成 $CS(A^T)^\perp=N(A)$。

如果要给 $CS(A)$ 也找一个正交补空间，那么它是 $N(A^T)$，我们叫它矩阵 A 的**左零空间（left null space）**。这样我们现在就有了关于矩阵 A 的四大基本子空间（four fundamental subspaces），它们分别在 $\mathbb{R}^n$ 和 $\mathbb{R}^m$ 中构成两对互为正交补的子空间：

$$\dim(\text{行空间}) + \dim(\text{零空间}) = n$$

$$\dim(\text{列空间}) + \dim(\text{左零空间}) = m$$

$$\dim(\text{行空间}) = \dim(\text{列空间}) = k$$

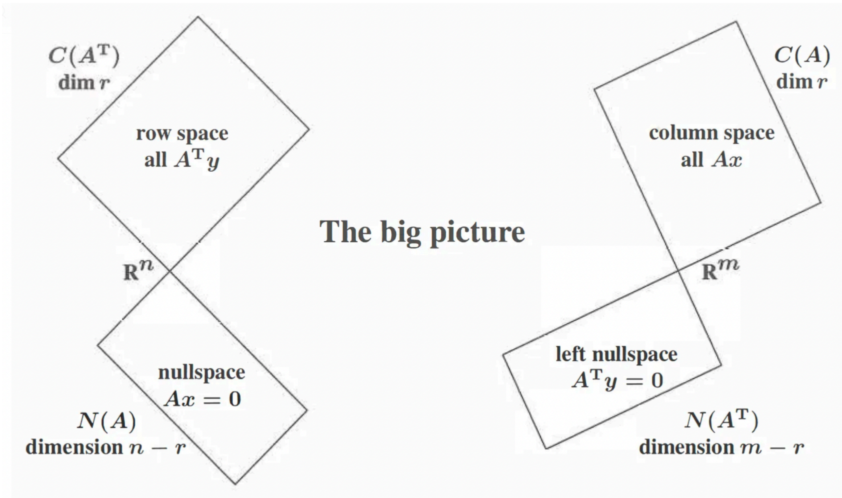

图 8: 矩阵的四个基本子空间

6.5 Exercises

7 Abstract Vector Spaces 抽象向量空间

7.1 Revisit Vector Spaces

在前面的章节中，我们详尽地探讨了定义在实数域上的 n 维向量空间 $\mathbb{R}^n$。我们熟悉其中的向量——它们是有序数组 $(x_1, x_2, \ldots, x_n)$，可以进行加法和数乘运算。我们对它的子空间、基和维数都已建立了牢固的几何与代数直观。

然而，$\mathbb{R}^n$ 仅仅是向量空间这个概念的一个最经典、最具体的模型。线性代数的力量与美感，很大程度上源于它所建立的理论框架具有惊人的普遍性。向量空间的公理化定义所描述的结构，在数学乃至科学和工程的无数领域中反复出现。

考虑以下这些集合，它们都与 $\mathbb{R}^n$ “看起来”完全不同，但都满足向量空间的公理：

- 所有次数不超过 n 的实系数多项式的集合 $\mathcal{P}_n$。
- 定义在区间 $[a, b]$ 上的所有连续实值函数的集合 $C([a, b])$。
- 所有 $m \times n$ 实矩阵的集合 $M_{m\times n}(\mathbb{R})$。
- 某个微分方程的所有解的集合。
- 复数域 $\mathbb{C}$（可视作实数域 $\mathbb{R}$ 上的向量空间）。

在 $\mathcal{P}_n$ 中，“向量”是多项式，如 $p(t) = a_0 + a_1 t + \cdots + a_n t^n$。两个多项式相加得到一个新的多项式，一个多项式乘以一个实数得到另一个多项式，零“向量”是零多项式。这与 $\mathbb{R}^{n+1}$ 在结构上同构（通过系数对应），但它本身的对象是函数。

在 $C([a, b])$ 中，“向量”是连续函数，如 $f(x) = \sin x$ 或 $g(x) = e^x$。函数的加法和数乘仍然产生连续函数。这个空间是无限维的，它没有有限个函数能张成整个空间。这已经远远超出了 $\mathbb{R}^n$ 的范畴。

当我们称这些结构为“向量空间”时，传统几何意义上“有方向有长度”的“向量”意象已经不再适用，甚至可能产生误导。一个函数或一个多项式，我们通常不将其想象为一个箭头。那么，这些迥异的对象为何共享同一个名字？

答案是：它们共享的是**运算的结构**，而非对象的具体形态。它们都装备了两种运算——加法和标量乘法，并且这些运算满足一系列确保其行为“线性”的公理（封闭性、结合律、交换律、零元、负元、分配律等）。这些公理的核心，正是我们在研究 $\mathbb{R}^n$ 时提炼出的**线性性**。

因此，一个更准确、更具包容性的名称是**线性空间**（linear space）。这个名称强调：

- **核心是线性结构**：关注的是对象之间如何通过加法和数乘进行“线性组合”，这是线性代数所有核心概念（线性相关/无关、基、维数、线性变换）的基石。
- **摆脱具体意象**：避免与几何向量捆绑，提醒我们空间中的元素可以是任何满足公理的事物——多项式、函数、矩阵、序列，甚至是更抽象的对象。
- **凸显普遍性**：直接点明这门学科（线性代数）的研究对象——具有线性运算结构的空间。

在本章中，为了更清晰地体现这一抽象观点，并避免因“向量”一词带来的不必要的具象化联想，我们将主要使用**线性空间**这一术语来指代满足向量空间公理的任何结构。我们仍会使用“向量”一词来指代线性空间中的单个元素，但请时刻记住，它可能是一个函数、一个矩阵，或任何其他东西。

在接下来的章节中，我们将看到，之前为 $\mathbb{R}^n$ 发展出的几乎所有概念——子空间、零空间、列空间、基、维数、坐标——都可以无缝地推广到抽象的线性空间上。这正是公理化方法的威力：一旦我们在一个具体模型（如 $\mathbb{R}^n$）中验证了某条定理仅依赖于公理，那么该定理在所有线性空间中自动成立。我们将探索这些抽象空间的例子，理解它们的基与维数，并学习如何将它们之间的线性变换与矩阵联系起来。

7.2 Examples of Abstract Vector Spaces

在上一节，我们列举了许多可能成为线性空间的集合。现在，让我们选取其中几个具体的例子，严格按照向量空间的十条公理进行验证。这种验证是理解“线性空间”作为公理化定义的本质所在。

例题 7.1 (多项式空间 $\mathcal{P}_n$) 令 $\mathcal{P}_n$ 表示所有次数不超过 n 的实系数多项式的集合，其中 n 是一个固定的非负整数。对于任意 $p(t), q(t) \in \mathcal{P}_n$ 和标量 $c \in \mathbb{R}$，我们定义：

$$(p+q)(t) := p(t) + q(t), \quad (cp)(t) := c \cdot p(t)$$

即，加法和数乘定义为函数的逐点加法和数乘。

我们需要验证 $\mathcal{P}_n$ 连同上述运算构成实数域 $\mathbb{R}$ 上的线性空间。设 $p(t) = a_0 + a_1 t + \cdots + a_n t^n$，$q(t) = b_0 + b_1 t + \cdots + b_n t^n$，以及 $r(t) = d_0 + d_1 t + \cdots + d_n t^n$。

1. **加法封闭性**：$p+q$ 的结果是 $(a_0+b_0) + (a_1+b_1)t + \cdots + (a_n+b_n)t^n$，这仍是一个次数不超过 n 的实系数多项式，故 $p+q \in \mathcal{P}_n$。
2. **标量乘法封闭性**：cp 的结果是 $ca_0 + (ca_1)t + \cdots + (ca_n)t^n$，也是一个次数不超过 n 的实系数多项式，故 $cp \in \mathcal{P}_n$。
3. **加法交换律**：$p+q = (a_0+b_0) + \cdots + (a_n+b_n)t^n = (b_0+a_0) + \cdots + (b_n+a_n)t^n = q+p$。
4. **加法结合律**：$(p+q)+r = p+(q+r)$ 由实数的加法结合律保证。
5. **零元存在性**：零多项式 $0(t) = 0$ 满足 $0 \in \mathcal{P}_n$，且对任意 $p \in \mathcal{P}_n$，有 $(p+0)(t) = p(t)+0 = p(t)$，故 0 是零元。
6. **负元存在性**：对于 $p(t)$，定义 $-p(t) := -a_0 - a_1 t - \cdots - a_n t^n$，则 $-p \in \mathcal{P}_n$，且 $(p+(-p))(t) = p(t) - p(t) = 0$，故 $-p$ 是 p 的负元。
7. **标量乘法单位元**：数字 $1 \in \mathbb{R}$ 满足 $(1p)(t) = 1 \cdot p(t) = p(t)$。
8. **标量乘法与域乘法相容**：对于 $c, d \in \mathbb{R}$，$((cd)p)(t) = (cd)p(t) = c(dp(t)) = (c(dp))(t)$。
9. **标量乘法对向量加法的分配律**：$c(p+q)(t) = c(p(t)+q(t)) = cp(t) + cq(t) = (cp+cq)(t)$。
10. **标量乘法对标量加法的分配律**：$(c+d)p(t) = cp(t) + dp(t) = (cp+dp)(t)$。

因此，$\mathcal{P}_n$ 是 $\mathbb{R}$ 上的线性空间。注意，它的“向量”是多项式，零向量是零多项式。

例题 7.2 (连续函数空间) 令 $C([a,b])$ 表示定义在闭区间 $[a,b]$ 上所有实值连续函数的集合。对于 $f, g \in C([a,b])$ 和 $c \in \mathbb{R}$，同样定义逐点运算：

$$(f+g)(x) := f(x) + g(x), \quad (cf)(x) := cf(x)$$

验证的关键在于利用连续函数的性质：连续函数的和与标量倍仍是连续函数。这直接保证了封闭性（公理 1, 2）。其余的公理（3-10）的验证与 $\mathcal{P}_n$ 完全类似，因为它们最终都归结为实数 $f(x), g(x)$ 在每一点 x 上满足相应的实数运算法则。例如：

- 零元是恒为零的函数 $0(x) \equiv 0$，它在 $[a, b]$ 上连续。
- f 的负元是 $(-f)(x) = -f(x)$。
- 所有分配律、结合律等均在每个 x 处由实数运算律保证。

因此，$C([a, b])$ 也是一个 $\mathbb{R}$ 上的线性空间。这个空间是**无限维**的。

例题 7.3 (矩阵空间 $M_{m\times n}(\mathbb{R})$) 所有 $m \times n$ 实矩阵的集合 $M_{m\times n}(\mathbb{R})$，在通常的矩阵加法和数乘下，构成线性空间。

验证是直接的：

- 封闭性：同型矩阵相加或数乘后仍是 $m \times n$ 矩阵。
- 零元：$m \times n$ 的零矩阵。
- 负元：矩阵 $A = [a_{ij}]$ 的负元是 $-A = [-a_{ij}]$。
- 其他公理由矩阵元素（实数）的运算律逐项保证。

这个线性空间与 $\mathbb{R}^{mn}$ 同构，维数为 mn。

7.3 Basis in Abstract Vector Spaces and Linear Isomorphism

在 $\mathbb{R}^n$ 中，基是一组线性无关且能张成整个空间的向量。这个核心概念可以完全平行地推广到任何线性空间。对于线性空间 V，一组向量 $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\} \subset V$ 被称为 V 的一个**基**，如果它满足以下两个条件：

1. **线性无关**：方程 $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$ 仅在标量 $c_1 = c_2 = \cdots = c_p = 0$ 时成立。
2. **张成空间**：V 中的每一个向量 $\mathbf{v}$ 都可以表示为 $\mathcal{B}$ 中向量的线性组合，即存在标量 $c_1, c_2, \ldots, c_p$ 使得 $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$。

如果 V 有一个由有限个向量构成的基，则称 V 是**有限维**的，且所有基包含的向量个数相同，这个数称为 V 的**维数**，记作 $\dim V$。否则，称 V 是**无限维**的。

现在，让我们在几个具体的线性空间中寻找和描述它们的基。

例题 7.4 (多项式空间 $\mathcal{P}_n$ 的标准基) 对于次数不超过 n 的多项式空间 $\mathcal{P}_n$，考虑以下 $n+1$ 个多项式：

$$\mathcal{B} = \{1, t, t^2, \ldots, t^n\}.$$

- **张成性**：任意多项式 $p(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n \in \mathcal{P}_n$ 显然可以写成 $\mathcal{B}$ 中元素的线性组合：$p = a_0 \cdot 1 + a_1 \cdot t + a_2 \cdot t^2 + \cdots + a_n \cdot t^n$。
- **线性无关性**：假设存在标量 $c_0, c_1, \ldots, c_n$ 使得

  $$c_0 \cdot 1 + c_1 \cdot t + c_2 \cdot t^2 + \cdots + c_n \cdot t^n = 0(t) \quad (\text{零多项式}).$$

  零多项式意味着对所有 $t \in \mathbb{R}$，该多项式值恒为零。根据多项式理论，一个多项式恒为零当且仅当其所有系数为零，故必有 $c_0 = c_1 = \cdots = c_n = 0$。因此 $\mathcal{B}$ 线性无关。

所以 $\mathcal{B}$ 是 $\mathcal{P}_n$ 的一个基，称为**标准基**。由此得到 $\dim \mathcal{P}_n = n+1$。注意，一个多项式的“坐标向量”正是它的系数向量 $(a_0, a_1, \ldots, a_n) \in \mathbb{R}^{n+1}$。

例题 7.5 (矩阵空间 $M_{m\times n}(\mathbb{R})$ 的标准基) 在 $m \times n$ 矩阵空间 $M_{m\times n}(\mathbb{R})$ 中，考虑以下矩阵：对于每个 $i = 1, \ldots, m$ 和 $j = 1, \ldots, n$，令 $\mathbf{E}_{ij}$ 表示一个 $m \times n$ 矩阵，其第 (i, j) 位置元素为 1，其余元素为 0。例如，在 $M_{2\times 3}(\mathbb{R})$ 中：

$$\mathbf{E}_{11} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{E}_{12} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \ldots, \quad \mathbf{E}_{23} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

所有这样的 $\mathbf{E}_{ij}$ 共有 mn 个。

- **张成性**：任意矩阵 $A = [a_{ij}] \in M_{m\times n}$ 可以唯一地表示为：

$$A = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \mathbf{E}_{ij}.$$

- **线性无关性**：假设存在标量 c_{ij} 使得 $\sum_{i,j} c_{ij} \mathbf{E}_{ij} = \mathbf{0}$（零矩阵）。这意味着得到的矩阵每个位置 (i, j) 的元素都是 c_{ij}，且必须等于零。因此所有 $c_{ij} = 0$。

所以，集合 $\{\mathbf{E}_{ij} \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ 是 $M_{m\times n}(\mathbb{R})$ 的一个基，称为**标准矩阵基**。维数为 $\dim M_{m\times n}(\mathbb{R}) = mn$。

例题 7.6 (对称矩阵子空间的基) 考虑 $n \times n$ 实对称矩阵的集合 $S_n = \{A \in M_{n\times n}(\mathbb{R}) \mid A^T = A\}$，它是 $M_{n\times n}(\mathbb{R})$ 的子空间。对称矩阵的特点是 $a_{ij} = a_{ji}$。我们可以为 S_n 构造一个基：

$$\mathcal{B}_{sym} = \{\mathbf{E}_{ii} \mid 1 \leq i \leq n\} \cup \{\mathbf{E}_{ij} + \mathbf{E}_{ji} \mid 1 \leq i < j \leq n\}.$$

- **解释**：基包含两类矩阵：
  1. 对角元：每个 $\mathbf{E}_{ii}$ 代表一个只在第 i 行第 i 列为 1 的矩阵，它显然是对称的。
  2. 非对角元：对于每对 $i < j$，矩阵 $\mathbf{E}_{ij} + \mathbf{E}_{ji}$ 在第 (i, j) 和 (j, i) 位置都是 1，其余为 0，也是对称的。
- **张成性**：任意对称矩阵 $A = [a_{ij}]$ 可以写成：

$$A = \sum_{i=1}^{n} a_{ii} \mathbf{E}_{ii} + \sum_{1 \leq i < j \leq n} a_{ij} (\mathbf{E}_{ij} + \mathbf{E}_{ji}).$$

- **线性无关性**：这些矩阵是线性无关的，因为它们来自 $M_{n\times n}$ 的标准基的特定线性组合，且彼此“支撑”在不同位置上。

基中向量的个数为：n *(*来自对角元*)* $+\frac{n(n-1)}{2}$ *(*来自上三角的 $i < j$ 对*)*。因此，

$$\dim S_n = n + \frac{n(n-1)}{2} = \frac{n(n+1)}{2}.$$

例题 7.7 (一个无限维空间的例子) 考虑连续函数空间 $C([a, b])$。我们断言它是无限维的。假设它是有限维的，维数为 N，并设 $\{f_1, f_2, \ldots, f_N\}$ 是一个基。那么任何连续函数，特别是幂函数 $1, x, x^2, \ldots, x^N$，都可以被这 N 个函数线性表示。这意味着存在不全为零的标量 $c_0, c_1, \ldots, c_N$ 使得函数 $g(x) = c_0 \cdot 1 + c_1 \cdot x + \cdots + c_N \cdot x^N$ 是零函数（因为 $N+1$ 个向量 $1, x, \ldots, x^N$ 在 N 维空间中必然线性相关）。但一个 N 次多项式除非所有系数为零，否则最多只有 N 个实根，不可能在整个区间 $[a, b]$ 上恒为零。这就产生了矛盾。因此，$C([a, b])$ 中不存在由有限个连续函数构成的基，它是**无限维**的。

一旦为线性空间 V 选定了一个基 $\mathcal{B}=\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$，$V$ 中的每一个向量 $\mathbf{v}$ 就唯一地对应一个坐标向量 $[\mathbf{v}]_{\mathcal{B}}=(c_1,c_2,\ldots,c_n)^T\in\mathbb{R}^n$，其中 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$。映射 $\mathbf{v}\mapsto[\mathbf{v}]_{\mathcal{B}}$ 是从 V 到 $\mathbb{R}^n$ 的一个**双射**，并且保持线性运算：

$$[\mathbf{u}+\mathbf{v}]_{\mathcal{B}}=[\mathbf{u}]_{\mathcal{B}}+[\mathbf{v}]_{\mathcal{B}},\quad [c\mathbf{v}]_{\mathcal{B}}=c[\mathbf{v}]_{\mathcal{B}}.$$

这样的映射称为（线性）**同构**。它告诉我们：**任何 n 维实线性空间 V，在选定一个基之后，本质上与 $\mathbb{R}^n$ 是"一样"的**。所有的线性关系、线性变换问题，都可以通过坐标转化到 $\mathbb{R}^n$ 中来处理。

- 对于 $\mathcal{P}_n$，取标准基 $\{1,t,\ldots,t^n\}$，多项式 $3-2t+5t^3$ 的坐标就是 $(3,-2,0,5,0,\ldots,0)^T\in\mathbb{R}^{n+1}$。
- 对于 $M_{2\times 2}$，取标准基 $\left\{\begin{bmatrix}1&0\\0&0\end{bmatrix},\begin{bmatrix}0&1\\0&0\end{bmatrix},\begin{bmatrix}0&0\\1&0\end{bmatrix},\begin{bmatrix}0&0\\0&1\end{bmatrix}\right\}$，矩阵 $\begin{bmatrix}2&-1\\3&4\end{bmatrix}$ 的坐标就是 $(2,-1,3,4)^T\in\mathbb{R}^4$。

7.4 Other Norms of $\mathbb{R}^n$

到目前为止，我们专注于线性空间的代数结构：加法和标量乘法。然而，在许多应用中（如分析收敛性、定义距离、优化问题），我们需要度量空间中向量的"大小"或向量之间的"角度"。这就是范数和内积的作用。它们为线性空间增添了额外的几何和分析结构。

在 $\mathbb{R}^n$ 中，我们最熟悉的是**欧几里得范数**（或 ℓ^2 范数）：

$$\|\mathbf{x}\|_2=\sqrt{x_1^2+x_2^2+\cdots+x_n^2}.$$

它源自**标准内积**（点积）$\langle\mathbf{x},\mathbf{y}\rangle=x_1y_1+x_2y_2+\cdots+x_ny_n$，满足 $\|\mathbf{x}\|_2=\sqrt{\langle\mathbf{x},\mathbf{x}\rangle}$。

然而，ℓ^2 范数只是 $\mathbb{R}^n$ 上众多可能的范数之一。其他常用的范数包括：

- **ℓ^1 范数（和范数）**：
$$\|\mathbf{x}\|_1=|x_1|+|x_2|+\cdots+|x_n|.$$
- **ℓ^∞ 范数（最大范数）**：
$$\|\mathbf{x}\|_\infty=\max\{|x_1|,|x_2|,\ldots,|x_n|\}.$$
- **更一般的 ℓ^p 范数（$1\le p<\infty$）**：
$$\|\mathbf{x}\|_p=\left(|x_1|^p+|x_2|^p+\cdots+|x_n|^p\right)^{1/p}.$$

  当 $p=2$ 时，即为欧几里得范数。

所有这些函数 $\|\cdot\|_p:\mathbb{R}^n\to[0,\infty)$ 都满足范数的四条公理：

1. **非负性**：$\|\mathbf{x}\|_p\ge 0$。
2. **确定性**：$\|\mathbf{x}\|_p=0 \iff \mathbf{x}=\mathbf{0}$。
3. **齐次性**：$\|c\mathbf{x}\|_p=|c|\|\mathbf{x}\|_p$，$c\in\mathbb{R}$。
4. **三角不等式**：$\|\mathbf{x}+\mathbf{y}\|_p\le\|\mathbf{x}\|_p+\|\mathbf{y}\|_p$。

它们定义了 $\mathbb{R}^n$ 上不同的"距离"概念。由不同范数诱导的距离 $d_p(\mathbf{x},\mathbf{y})=\|\mathbf{x}-\mathbf{y}\|_p$ 在几何上表现为不同的"单位球"形状：

- ℓ^1：单位球是菱形（在 $\mathbb{R}^2$ 中是正方形旋转 $45°$）。
- ℓ^2：单位球是标准的圆形/球体。
- ℓ^∞：单位球是正方形/立方体。

何时使用不同的范数？

- ℓ^2 **范数**：最常见，具有旋转不变性，对应于最小二乘法、物理学中的能量。
- ℓ^1 **范数**：在统计学（Lasso 回归）、信号处理、稀疏表示中至关重要，因为它倾向于产生稀疏解（许多分量为零）。
- ℓ^∞ **范数**：关注的是最大误差，在控制理论、worst-case 分析中常用。
- **一般 ℓ^p 范数**：在泛函分析、插值理论中出现。

尽管这些范数定义不同，但由于 $\mathbb{R}^n$ 是有限维的，它们在拓扑意义上是**等价**的：存在常数 $c, C > 0$ 使得 $c\|\mathbf{x}\|_p \le \|\mathbf{x}\|_q \le C\|\mathbf{x}\|_p$ 对所有 $\mathbf{x}$ 成立。这意味着序列的收敛性在所有范数下是一致的。但在无限维空间中，这种等价性不再成立。

7.5 Norms and Inner Products of Abstract Vector Spaces

现在，我们将这些概念推广到一般的线性空间 V。

定义 7.1 (赋范空间) 设 V 是域 $\mathbb{R}$（或 $\mathbb{C}$）上的线性空间。一个函数 $\|\cdot\| : V \to [0, \infty)$ 称为 V 上的一个**范数**，如果它满足前述四条公理。二元组 $(V, \|\cdot\|)$ 称为一个**赋范线性空间**。

定义 7.2 (内积空间) 设 V 是域 $\mathbb{R}$（或 $\mathbb{C}$）上的线性空间。一个函数 $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$（或 $\mathbb{C}$）称为 V 上的一个**内积**，如果它满足：

1. **对称性（或共轭对称性）**：$\langle \mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{v}, \mathbf{u}\rangle$（在复数域上为 $\overline{\langle \mathbf{v}, \mathbf{u}\rangle}$）。
2. **线性性（对第一个变量）**：$\langle a\mathbf{u} + b\mathbf{v}, \mathbf{w}\rangle = a\langle \mathbf{u}, \mathbf{w}\rangle + b\langle \mathbf{v}, \mathbf{w}\rangle$。
3. **正定性**：$\langle \mathbf{v}, \mathbf{v}\rangle \ge 0$，且 $\langle \mathbf{v}, \mathbf{v}\rangle = 0 \iff \mathbf{v} = \mathbf{0}$。

二元组 $(V, \langle\cdot,\cdot\rangle)$ 称为一个**内积空间**。在内积空间中，可以**由内积诱导一个范数**：

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}.$$

这个范数自动满足三角不等式（由柯西-施瓦茨不等式 $|\langle \mathbf{u}, \mathbf{v}\rangle| \le \|\mathbf{u}\|\|\mathbf{v}\|$ 保证）。因此，每个内积空间自然是一个赋范空间。

现在，让我们为之前讨论的抽象线性空间装备上常见的范数和内积。

1. 多项式空间 $\mathcal{P}_n$

我们可以通过多种方式将 $\mathcal{P}_n$ 视为赋范空间。

- **系数范数**：利用 $\mathcal{P}_n$ 与 $\mathbb{R}^{n+1}$ 的同构（通过标准基 $\{1, t, \ldots, t^n\}$），我们可以将 $\mathbb{R}^{n+1}$ 的任何范数"拉回"到 $\mathcal{P}_n$ 上。例如，对于 $p(t) = a_0 + a_1 t + \cdots + a_n t^n$，定义：

  $$\|p\|_1 = |a_0| + |a_1| + \cdots + |a_n|, \quad \|p\|_2 = \sqrt{a_0^2 + a_1^2 + \cdots + a_n^2}, \quad \|p\|_\infty = \max\{|a_0|, |a_1|, \ldots, |a_n|\}.$$

  这些是 $\mathcal{P}_n$ 上完全合法的范数，但它们的几何意义是基于系数的，而非多项式函数本身的性质。

- **函数范数**：由于多项式是函数，我们可以基于函数在区间上的值来定义范数。例如，给定区间 $[a,b]$，定义：

$$\|p\|_{C([a,b])} = \max_{t\in[a,b]} |p(t)| \quad (\text{一致范数或上确界范数}),$$

$$\|p\|_{L^2([a,b])} = \left(\int_a^b |p(t)|^2\,dt\right)^{1/2} \quad (L^2\text{范数}).$$

这些范数捕捉了多项式作为函数的整体行为，与系数范数可能给出完全不同的“大小”度量。验证三角不等式等需要用到实分析的工具。

- **内积**：在 $\mathcal{P}_n$ 上可以定义内积，从而诱导范数。例如，L^2 内积：

$$\langle p,q\rangle_{L^2} = \int_a^b p(t)q(t)\,dt.$$

容易验证它满足内积公理。由此诱导的范数正是 $\|p\|_{L^2([a,b])}$。注意，同一个线性空间 $\mathcal{P}_n$，装备不同的内积（如对应不同区间或不同权重函数）会成为不同的内积空间。

2. 连续函数空间 $C([a,b])$

这是一个无限维空间，范数和内积的选择更为关键。

- **一致范数（上确界范数）**：

$$\|f\|_\infty = \sup_{t\in[a,b]} |f(t)| = \max_{t\in[a,b]} |f(t)| \quad (\text{由于连续性，最大值存在}).$$

这是 $C([a,b])$ 上最重要和自然的范数之一。在此范数下，函数序列 f_n 收敛到 f 意味着 f_n 在 $[a,b]$ 上**一致收敛**于 f。装备此范数的 $C([a,b])$ 是一个完备的赋范空间（巴拿赫空间）。

- **L^p 范数**（$1\le p<\infty$）：

$$\|f\|_{L^p} = \left(\int_a^b |f(t)|^p\,dt\right)^{1/p}.$$

当 $p=2$ 时，我们有 L^2 范数。注意，按照此定义，若 $\|f\|_{L^p}=0$，只能推出 f **几乎处处**为零，对于连续函数，这意味着 f 恒为零。因此，在 $C([a,b])$ 上，$\|\cdot\|_{L^p}$ 确实是一个范数。

- **L^2 内积**：

$$\langle f,g\rangle_{L^2} = \int_a^b f(t)g(t)\,dt.$$

这使得 $C([a,b])$ 成为一个内积空间（虽然不完备）。由此内积可以定义函数的“正交性”，这是傅里叶级数理论的起点：三角函数系 $\{1,\cos(nt),\sin(nt)\}$ 在 $L^2([-\pi,\pi])$ 内积下是正交的。

3. 矩阵空间 $M_{m\times n}(\mathbb{R})$

- **Frobenius 范数（希尔伯特-施密特范数）**：将矩阵 A 视为 mn 维向量，应用欧几里得范数：

$$\|A\|_F = \sqrt{\sum_{i=1}^m\sum_{j=1}^n |a_{ij}|^2} = \sqrt{\operatorname{tr}(A^TA)}.$$

这是一个非常常用的范数，易于计算，且由以下内积诱导：

$$\langle A,B\rangle_F = \operatorname{tr}(B^TA) = \sum_{i,j} a_{ij}b_{ij}.$$

- **算子范数（由向量范数诱导）**：将矩阵视为线性算子 $A:\mathbb{R}^n\to\mathbb{R}^m$，其范数定义为它能将向量“拉伸”的最大倍数：
$$\|A\|_{\text{op}}=\sup_{\mathbf{x}\neq 0}\frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}=\sup_{\|\mathbf{x}\|=1}\|A\mathbf{x}\|.$$
根据所选的向量范数（$\ell^1,\ell^2,\ell^\infty$），可以得到不同的算子范数：
  - $\|A\|_1$（列和范数）：各列绝对值之和的最大值。
  - $\|A\|_2$（谱范数）：A 的最大奇异值，$\sqrt{\lambda_{\max}(A^TA)}$。
  - $\|A\|_\infty$（行和范数）：各行绝对值之和的最大值。

  这些算子范数满足次可乘性 $\|AB\|\le\|A\|\|B\|$，在数值分析和线性系统理论中至关重要。

- **核范数**：
$$\|A\|_*=\sum_i\sigma_i,\quad(\sigma_i\text{为 }A\text{ 的奇异值}).$$
这不是由内积诱导的，但在低秩矩阵恢复和压缩感知中极其重要。

4. 平方可和序列空间 ℓ^2

定义 7.3 (ℓ^2 空间) 定义 ℓ^2 为所有平方可和的实（或复）序列的集合：
$$\ell^2=\left\{\mathbf{x}=(x_1,x_2,x_3,\dots):x_i\in\mathbb{R}\ (\text{或 }\mathbb{C}),\ \sum_{i=1}^{\infty}|x_i|^2<\infty\right\}.$$
即，序列元素的平方和收敛。

- **线性空间结构**：ℓ^2 在逐项加法和数乘下构成线性空间。封闭性的关键在于闵可夫斯基不等式：
$$\left(\sum_{i=1}^{\infty}|x_i+y_i|^2\right)^{1/2}\le\left(\sum_{i=1}^{\infty}|x_i|^2\right)^{1/2}+\left(\sum_{i=1}^{\infty}|y_i|^2\right)^{1/2}.$$

- **标准内积**：对于 $\mathbf{x},\mathbf{y}\in\ell^2$，定义内积：
$$\langle\mathbf{x},\mathbf{y}\rangle=\sum_{i=1}^{\infty}x_i\overline{y_i}\quad(\text{实数域时去掉共轭}).$$
由柯西-施瓦茨不等式保证该级数绝对收敛：
$$\left|\sum_{i=1}^{\infty}x_i\overline{y_i}\right|\le\left(\sum_{i=1}^{\infty}|x_i|^2\right)^{1/2}\left(\sum_{i=1}^{\infty}|y_i|^2\right)^{1/2}.$$

- **诱导的 ℓ^2 范数**：
$$\|\mathbf{x}\|_2=\sqrt{\langle\mathbf{x},\mathbf{x}\rangle}=\left(\sum_{i=1}^{\infty}|x_i|^2\right)^{1/2}.$$
装备此范数的 ℓ^2 空间是完备的（任何柯西序列都收敛于 ℓ^2 中的元素）。

- **其他范数**：类似于 $\mathbb{R}^n$，可以定义 ℓ^p 空间 $(1\le p\le\infty)$：
$$\ell^p=\left\{\mathbf{x}:\|\mathbf{x}\|_p=\left(\sum_{i=1}^{\infty}|x_i|^p\right)^{1/p}<\infty\right\},\quad 1\le p<\infty,$$
$$\ell^\infty=\left\{\mathbf{x}:\|\mathbf{x}\|_\infty=\sup_{i\ge1}|x_i|<\infty\right\}.$$
注意：ℓ^p 空间对不同的 p 是不同的集合。例如，序列 $(1,\frac{1}{2},\frac{1}{3},\dots)$ 属于 ℓ^2 但不属于 ℓ^1。

为抽象线性空间装备范数或内积，使其从纯粹的代数对象升级为具有几何和分析结构的对象：

- **赋范空间**：允许我们谈论向量的长度、序列的收敛性、连续性、以及空间的完备性（巴拿赫空间）。
- **内积空间**：更进一步，允许我们谈论角度、正交性、投影（如最小二乘逼近）。完备的内积空间称为希尔伯特空间，是泛函分析的核心。

同一个线性空间可以装备多种不同的范数和内积，选择哪一种取决于具体问题的物理背景或数学需求。例如，在多项式逼近中，若关心最大误差，则用一致范数；若关心平均误差，则用 L^2 范数。在数据科学中，矩阵的 Frobenius 范数常用于正则化，而谱范数则与奇异值分解和稳定性分析相关。

理解这些结构，是迈向泛函分析、数值分析、信号处理以及现代机器学习理论的重要一步。

表 2: 常见线性空间上的内积、范数与距离

空间	内积与诱导范数	其他常见范数	诱导的距离
$\mathbb{R}^n$	**内积:** $\langle \mathbf{x}, \mathbf{y} \rangle = \sum x_i y_i$ **诱导范数:** $\|\mathbf{x}\|_2 = \sqrt{\sum x_i^2}$	$\|\mathbf{x}\|_1 = \sum \|x_i\|$ $\|\mathbf{x}\|_\infty = \max \|x_i\|$ $\|\mathbf{x}\|_p = (\sum \|x_i\|^p)^{1/p}$	$d_2(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_2$ $d_1(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_1$ $d_\infty(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_\infty$ $d_p(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_p$
$C([a,b])$	**内积:** $\langle f, g \rangle_{L^2} = \int_a^b fg\,dt$ **诱导范数:** $\|f\|_{L^2} = \sqrt{\int_a^b f^2\,dt}$	$\|f\|_\infty = \max\limits_{t\in[a,b]} \|f(t)\|$ $\|f\|_{L^p} = (\int_a^b \|f\|^p\,dt)^{1/p}$	$d_{L^2}(f,g) = \|f-g\|_{L^2}$ $d_\infty(f,g) = \|f-g\|_\infty$ $d_{L^p}(f,g) = \|f-g\|_{L^p}$
$\mathcal{P}_n$	**系数内积:** $\langle p, q \rangle_c = \mathbf{a}^T\mathbf{b}$ **诱导范数:** $\|p\|_{c,2} = \sqrt{\sum a_i^2}$ L^2 **内积:** $\langle p, q \rangle_{L^2} = \int_a^b pq\,dt$ **诱导范数:** $\|p\|_{L^2} = \sqrt{\int_a^b p^2\,dt}$	$\|p\|_{c,1} = \sum \|a_i\|$ $\|p\|_{c,\infty} = \max \|a_i\|$ $\|p\|_\infty = \max\limits_{t\in[a,b]} \|p(t)\|$	$d_{c,2}(p,q) = \|\mathbf{a}-\mathbf{b}\|_2$ $d_{L^2}(p,q) = \|p-q\|_{L^2}$ $d_\infty(p,q) = \|p-q\|_\infty$
$M_{m\times n}$	**Frobenius 内积:** $\langle A, B \rangle_F = \mathrm{tr}(B^T A)$ **诱导范数:** $\|A\|_F = \sqrt{\sum a_{ij}^2}$	**算子范数:** $\|A\|_p = \sup\limits_{\mathbf{x}\neq 0} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$ 特例: $\|A\|_1$ (列和) $\|A\|_2$ (谱范数) $\|A\|_\infty$ (行和) **核范数:** $\|A\|_* = \sum \sigma_i(A)$	$d_F(A,B) = \|A-B\|_F$ $d_{\mathrm{op},p}(A,B) = \|A-B\|_p$
ℓ^2	**内积:** $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^\infty x_i y_i$ **诱导范数:** $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^\infty x_i^2}$	$\|\mathbf{x}\|_p = (\sum_{i=1}^\infty \|x_i\|^p)^{1/p}$ $\|\mathbf{x}\|_\infty = \sup\limits_i \|x_i\|$	$d_2(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_2$ $d_p(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_p$

1. **诱导关系链：** 内积 ⇒ 范数 ⇒ 距离。即：

$$\text{内积}\langle\cdot,\cdot\rangle \;\Rightarrow\; \text{范数}\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle} \;\Rightarrow\; \text{距离}d(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\|.$$

   反之不一定成立。

2. **有限维 vs 无限维：**
   - **有限维**（$\mathbb{R}^n$, $\mathcal{P}_n$, $M_{m\times n}$）：所有范数等价，空间完备。
   - **无限维**（$C([a,b])$, ℓ^2）：不同范数可能不等价，完备性依赖范数选择。

3. **典型应用：**

- ℓ^1：稀疏优化；ℓ^2：最小二乘；ℓ^∞：最坏情况分析。
- $\|\cdot\|_\infty$：一致逼近；$\|\cdot\|_{L^2}$：傅里叶分析。
- $\|\cdot\|_F$：矩阵正则化；$\|\cdot\|_2$（谱）：稳定性；$\|\cdot\|_*$：低秩恢复。

7.6 Exercises

8 Linear Transformations I 线性变换 I（基本概念）

矩阵可以建构空间这个功能有什么用呢？其实这都是在为本章要介绍的矩阵的第三个功能做铺垫，这也是矩阵最重要最终极的一个身份：它代表一个线性变换。

8.1 Linear Transformations: functions between vector spaces

线性变换是向量空间之间的一类特殊函数。

设 V 和 W 是向量空间。一个函数 $T : V \to W$ 被称为**线性变换 (linear transformation)**，如果对于所有向量 $\mathbf{u}, \mathbf{v} \in V$ 和所有标量 $\alpha \in \mathbb{R}$，满足：

(1) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$

(2) $T(\alpha\mathbf{u}) = \alpha T(\mathbf{u})$

当 $W = V$ 时，我们称线性变换 $T : V \to V$ 为一个**线性算子（linear operator）**。当 $W = \mathbb{F}$ 时，T 被称为**线性泛函（linear functional）**。线性泛函是将向量映射到标量的线性映射。

线性变换是**保持 (preserve)** 加法和数乘的函数。这意味着线性变换不会破坏向量空间的线性结构。请从代数和几何两个层面理解这句话：代数角度而言，向量空间的线性结构是指向量间的线性组合，而原本通过线性组合构造出来的向量在变换后依然是同样比例的组合。几何角度，你要对线性变换是一类什么样的变换有清晰的了解，因为它会影响你在之后判断一个场景是否可以应用线性代数的知识。想象现在把一个正方形施加一个线性变换，它可能变成平行四边形、瘪瘪的菱形，但边仍然平行，角的相对结构还在，原点仍在原点（不会漂移），所以它"看起来扁了"，但"结构没坏"。这就是线性变换保留结构的几何意义。之后我们会讨论线性变换可以做哪些变换（比如旋转？拉伸？翻转？平移?）。

- **恒等变换 (identity transformation)**: $T(\mathbf{v}) = \mathbf{v}$
- 两个线性变换的**复合 (composition)** 仍然是线性变换: $(S \circ T)(\mathbf{v}) = (ST)(\mathbf{v}) = S(T(\mathbf{v}))$
- 线性变换的线性组合仍然是线性变换: $S + T$, αS
- 如果存在，线性变换 T 的**逆 (inverse)** T^{-1} 是唯一的线性变换，满足: $T^{-1}(T(\mathbf{v})) = \mathbf{v}$

假设 T 是一个从向量空间 V 到向量空间 W 的线性变换。

线性变换 T 的**值域 (Range)**，记为 $R(T)$，定义为：

$$R(T) = \{\mathbf{w} \in W \mid \mathbf{w} = T(\mathbf{v}) \text{ 对于某个} \mathbf{v} \in V\} \subseteq W.$$

值域是 W 的一个子集，包含了所有能被 T 映射到的向量。

线性变换 T 的**核 (Kernel)**，记为 $\ker(T)$，定义为：

$$\ker(T) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}\} \subseteq V,$$

其中 $\mathbf{0}$ 是 W 的零向量。核是 V 的一个子集，包含了所有被 T 映射到 W 的零向量的向量。

定理 8.1 线性变换 $T : V \to W$ 的核 $\ker(T)$ 和值域 $R(T)$ 分别是 V 和 W 的子空间。

8.2 Linear Operator

在上节我们说了线性算子是当线性变换从一个向量空间映射到它自身时的特例。本节将介绍一些常见的线性算子，我们假设运算都是在合适的函数空间或向量空间上定义的。其线性性读者可自行验证。

1. **微分算子**：$D = \frac{d}{dx}$。它将一个可微函数 $f(x)$ 映射为其导数 $f'(x)$。
2. **积分算子**：$(If)(x) = \int_a^x f(t)dt$。它将函数 f 映射为其从 a 到 x 的变上限积分结果。
3. **矩阵算子**（在有限维空间）：对于一个 $n \times n$ 方阵 A，定义算子 $T : \mathbb{R}^n \to \mathbb{R}^n$ 为 $T(\mathbf{x}) = A\mathbf{x}$。它将一个 n 维向量映射为另一个 n 维向量。几何上，它可以表示旋转、缩放、剪切等变换。
4. **拉普拉斯算子**：$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$。它将一个二元函数 $f(x, y)$ 映射为其二阶偏导数的和。
5. 恒等算子与零算子：这两个是平凡但重要的例子。
   - **恒等算子** I：将每个元素映射为自身，即 $I(v) = v$。
   - **零算子** 0：将每个元素映射为零向量，即 $0(v) = 0$。

8.3 Linear VS non-Linear

本节回答一组自然而然的疑问：我们也许知道线性变换的应用场景很广，很多变换都可以看作是线性变换，继而可以应用线性代数的知识，比如图片处理，但哪些场景是非线性变换继而线性代数知识也许无能为力的？这也引出了另一个更大的问题，就是为什么数学里只有线性代数，但没有一个领域叫非线性代数？难道是非线性的情况都可以通过某些手段转换为线性问题吗？如果是的话这是全部原因吗？如果只是部分可以转换的话，那么哪些非线性情况无法转换为线性问题呢？

我们也许可以说，线性结构是我们能彻底系统性理解并驾驭的极限，而非线性是复杂性的本源。我们分四个问题来讨论：

首先，**什么是非线性变换？**非线性变换就是不满足以上两条定义的变换，常见的非线性变换例子包括：所有带有乘法、指数、对数、三角函数的映射，图像处理中的直方图均衡化（非线性映射灰度），物理系统中的混沌行为，神经网络的激活函数，非线性微分方程描述的系统（如气候、人口、金融市场），动态系统中具有“阈值效应”或“爆炸式增长”的行为。这些问题中，线性代数很多时候无能为力。

第二，**为什么线性结构特别？**为什么线性代数能成为一个独立学科？有几个主要原因：

1. 线性结构是稳定的：
   - 一组线性变换组合起来仍然是线性变换
   - 线性系统有“解的叠加性”，极大简化理解与求解
   - 可表示为矩阵，计算性极强（线性代数 = 数学 + 算法）
2. 所有空间的局部都近似是线性的：
   - 线性变换是“理解任意变换的第一步”

第三，**非线性问题能不能转为线性问题？**哪些能，哪些不能？

很多非线性问题都可以在某种条件下被线性近似或线性处理，其中一个主要思想是：很多非线性的对象，在合适的无限维空间中，可以通过无限多个线性基底逼近或展开。但值得注意的是，表达能力不等于结构本身是线性的。非线性系统之所以复杂，往往正是因为它们不满足线性叠加原则，即使能在更大的空间中被线性展开。几个经典例子包括：泰勒展开、傅立叶级数、神经网络。当然，也有些系统是根本不能被有效线性化的，比如：混沌系统、高度耦合非线性系统、非连续性突变行为。

第四，**为什么没有“非线性代数”这个学科分支？**

实际上有，只是叫法不同：如果你理解“代数”是处理方程与结构的学科，那么“非线性代数”其实是分散在多个分支中的。

实际分支名称	**本质上处理的是非线性结构/方程**
代数几何	解非线性多项式方程组（代数曲线、曲面）
非线性分析	非线性微分方程、泛函分析等
代数拓扑	用代数方法处理非线性几何形状
非线性动力系统	分岔、混沌、周期行为
微分几何	曲率、流形（不再是平的空间）

所以，“非线性代数”不是没有，而是它太复杂，必须分拆为多个强大的领域分别处理。

线性代数之所以成为独立的核心数学分支，是因为线性结构具备高度的稳定性、可解性和可计算性。非线性问题虽然更加广泛，但缺乏统一结构，通常只能局部近似为线性来处理。实际上，许多数学分支（如代数几何、微分几何、非线性动力系统）正是“非线性代数”的不同面貌，它们共同面对线性代数力所不及的复杂现象。

8.4 Matrix as a Linear Transformation

从这节开始我们关注有限维向量空间之间的线性变换。

本节核心目标是证明：**任意一个有限维向量空间之间的线性变换，都可以用一个矩阵来表示**，而变换作用等价于左乘该矩阵。这奠定了线性代数中几何（变换）与代数（矩阵）统一的基石。

容易看出，对于一个 $m \times n$ 矩阵 A，公式 $T(\mathbf{v}) = A\mathbf{v}$ 定义了一个线性变换 $T : \mathbb{R}^n \to \mathbb{R}^m$。现在，我们要探讨其逆命题：**对于任意一个线性变换 $T : \mathbb{R}^n \to \mathbb{R}^m$，是否都存在一个矩阵 A_T，使得 $T(\mathbf{v}) = A_T\mathbf{v}$？**答案是肯定的。

定理 8.2 ($\mathbb{R}^n \to \mathbb{R}^m$ 线性变换的矩阵表示) *设 $T : \mathbb{R}^n \to \mathbb{R}^m$ 是一个线性变换。令 $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ 是 $\mathbb{R}^n$ 的标准基。构造矩阵 A_T，使其第 i 列恰好是向量 $T(\mathbf{e}_i) \in \mathbb{R}^m$，即：*

$$A_T = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \end{bmatrix}.$$

那么，对于任意 $\mathbf{x} \in \mathbb{R}^n$，都有：

$$T(\mathbf{x}) = A_T\mathbf{x}.$$

理解这一构造的关键在于认识到：**一个线性变换由它作用于一组基向量的结果完全确定**。

- 对于 $\mathbb{R}^n$ 中的任意向量 $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n$。
- 根据线性变换的性质：$T(\mathbf{x}) = x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \cdots + x_nT(\mathbf{e}_n)$。
- 观察右边，这正好是矩阵 A_T 的列向量以 $x_1, x_2, \ldots, x_n$ 为系数的线性组合，根据矩阵乘法的定义，这恰恰就是 $A_T\mathbf{x}$。

因此，通过将标准基的像作为列向量，我们得到的矩阵 A_T 完美地“编码”了整个线性变换 T 的信息。

易看出，线性变换 $T(\mathbf{x}) = A\mathbf{x}$ 的值域和核与矩阵 A 的子空间存在着如下对应关系：

$$R(T) = CS(A), \quad \ker(T) = N(A).$$

现在我们可以从线性变换的角度来重新理解秩-零化度定理。

定理 8.3 (线性变换的秩-零化度定理) 设 T 是从有限维向量空间 V 到向量空间 W 的线性变换，则：

$$\mathrm{rank}(T) + \mathrm{nullity}(T) = \dim(V)$$

其中：

- $\mathrm{rank}(T) = \dim(R(T))$ 是变换的秩，即值域的维度
- $\mathrm{nullity}(T) = \dim(\ker(T))$ 是零化度，即核的维度

秩-零化度定理揭示了线性变换中信息流动的守恒关系：

- $\dim(V)$：输入信息的总量
- $\mathrm{nullity}(T)$：在变换过程中被丢弃或压缩为零的信息量
- $\mathrm{rank}(T)$：经过变换后保留下来的有效信息量

定理断言：有效信息量与被丢弃的信息量之和等于原始信息总量。

8.5 Geometric Interpretation of Linear Transformations

线性变换的几何含义是：**对向量空间进行一种“保持网格线平行且等距”的形变**。

线性变换是从 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的映射，那么其中一类特殊的线性变换就是当 $m = n$ 时，变换并未改变原本空间的维度。可以想到这类变换包括旋转、反射、伸缩等，其对应的矩阵 A 是 $n \times n$ 的方阵。这类变换比较简单，它们在 $\mathbb{R}^2$ 或 $\mathbb{R}^3$ 中的矩阵表示都有标准形式，我们稍后会具体来讨论。

- **旋转 (Rotation)**：平面绕原点旋转角度 θ。
- **反射 (Reflection)**：关于某条过原点的直线或某个过原点的平面的镜像。
- **伸缩 (Stretch/Scaling)**：沿坐标轴方向的拉伸或压缩。

很多初学者对于线性变换的几何想象就停留在这些不改变空间维度的基础变换了，这是极大的误解。线性变换完全可以改变空间维度，事实上，这才是更一般的情况。

1. 升维变换（$n < m$，A 是“瘦高”矩阵）
   - 特殊例子：嵌入变换 $T : \mathbb{R}^2 \to \mathbb{R}^3$，定义为 $T\left(\begin{pmatrix} x & y \end{pmatrix}\right) = \begin{pmatrix} x & y & 0 \end{pmatrix}$。
   - 几何意义：将二维平面“抬升”并放入三维空间中。
2. 降维变换（$n > m$，A 是“矮胖”矩阵）
   - 特殊例子：投影变换 $T : \mathbb{R}^3 \to \mathbb{R}^2$，定义为 $T\left(\begin{pmatrix} x & y & z \end{pmatrix}\right) = \begin{pmatrix} x & y \end{pmatrix}$。
   - 几何意义：将三维空间中的点“拍扁”到 xy-平面上，丢弃 z 坐标信息。

注意，当我们说一个变换 $T : \mathbb{R}^n \to \mathbb{R}^m$ 是“升维的”，通常指它的**上域（Codomain）**维度 m 大于其**定义域（Domain）**维度 n。虽然输出生活在更高维的空间 $\mathbb{R}^m$ 中，但变换实际的输出结果——即**值域（Range）** $R(T)$——只是这个高维空间里的一个子集，并且这个子集的维度**不可能超过**输入空间的维度。

根据秩-零化度定理：

$$\underbrace{\dim(R(T))}_{\text{值域维度}} = \underbrace{\dim(V)}_{\text{输入维度}} - \underbrace{\dim(\ker(T))}_{\text{核的维度}} \leq \dim(V)$$

因此，对于一个 $m \times n$ 矩阵 A（$m > n$）：

- 它确实将向量从 $\mathbb{R}^n$ **映射到** $\mathbb{R}^m$。
- 但所有可能的输出向量 $A\mathbf{x}$ 都位于 $\mathbb{R}^m$ 中的一个**至多 n 维**的子空间里。
- 这个子空间可能是 $\mathbb{R}^m$ 中的一个平面、直线，甚至一个点。

考虑嵌入变换 $T : \mathbb{R}^2 \to \mathbb{R}^3$，定义为：

$$T\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$$

其矩阵为 $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$。

- **外观**：我们将二维平面上的点 (x, y) 送到了三维空间。
- **实质**：所有输出点都落在三维空间的 xy-平面上。这个平面本身是一个二维子空间。
- **值域维度**：$\dim(R(T)) = 2$，等于输入维度 $\dim(\mathbb{R}^2) = 2$。
- **并未真正"升维"**：变换没有创造出新的独立方向（z 坐标始终为 0）。

这是一个很直观的结论，从信息量角度考虑，我可以通过损失部分信息实现降维，但我没法凭空增加信息实现真正意义上的升维。我可以把背景空间的维度一直增加，但真正存储信息的那部分子空间的维度是不可能被升高的。

理解了线性变换的几何含义，我们可以回过头来回答一个问题。在之前我们说"如果存在，线性变换的逆是唯一的线性变换"，很多人可能会有疑惑：所以一个线性变换的逆变换有可能不存在？这似乎是反直觉的，因为如果你能完成一个变换过程，将这个过程逆着做一遍似乎是没有阻碍的。事实上，即使独立的升维变换和降维变换是存在的，任何一个升维或降维变换都是不存在其对应的逆变换的，也就是说改变空间维度的变换是不可逆变换。让我们顺势来讨论一下线性变换存在逆变换的充要条件。

一个线性变换 $T : V \to W$ 的逆变换 $T^{-1} : W \to V$ 存在的**充要条件**是：T 是一个**双射**（既单射又满射）。

一个直接的推论就是维度守恒是一个重要的必要条件，即可逆线性变换必须发生在两个维数相同的空间之间。

$$\dim(V) = \dim(W)$$

其矩阵表示 A 必须是方阵。

若 $\dim(V) < \dim(W)$，则 T 不可能是满射。值域 $R(T)$ 的维度最多为 $\dim(V)$，无法覆盖整个 W，因此存在向量没有原像。若 $\dim(V) > \dim(W)$，则 T 不可能是单射。根据秩-零化度定理，核 $\ker(T)$ 的维度至少为 $\dim(V) - \dim(W) > 0$，因此存在多个不同的向量被映射到同一个点。

当然，即使满足 $\dim(V) = \dim(W)$，逆变换也可能不存在。还需要满足单射或者满射（已知维度相同时，两者是等价的）。

考虑 $T : \mathbb{R}^2 \to \mathbb{R}^2$，将平面投影到 x 轴上：

$$T\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x \\ 0 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

显然，该变换不是单射，所有形如 $\begin{pmatrix} x \\ y_1 \end{pmatrix}$ 和 $\begin{pmatrix} x \\ y_2 \end{pmatrix}$ 的向量 $(y_1 \neq y_2)$ 都被映射到同一个点 $\begin{pmatrix} x \\ 0 \end{pmatrix}$。因此该变换不可逆，即便它不改变空间维度。从几何角度看，整个二维平面被“压扁”到一维的 x 轴上，y 坐标的信息丢失，无法从输出唯一确定输入。

总结而言，可逆的线性变换是一个“无损”的重新排列，它不压缩空间，也不丢失任何信息。其本质是对空间进行旋转、反射、伸缩等基本操作的组合。

对于有限维向量空间之间的线性变换，其可逆性与矩阵可逆性之间存在完美的对应。设 $T : V \to W$ 是线性变换，$\dim(V) = \dim(W) = n$。选定 V 的基 B 和 W 的基 C 后，T 有矩阵表示 $[T]_{C\leftarrow B} = A$（一个 $n \times n$ 矩阵）。则：

$$\boxed{T \text{ 可逆} \quad \Longleftrightarrow \quad A \text{ 可逆}}$$

并且，逆变换 T^{-1} 的矩阵表示就是 A^{-1}，即 $[T^{-1}]_{B\leftarrow C} = A^{-1}$。

从这点也可以再次验证改变维度的线性变换是不可逆的，因为其对应的矩阵不是方阵，自然不存在逆矩阵。

当 $T : \mathbb{R}^n \to \mathbb{R}^n$ 由方阵 A 表示时，以下陈述全部等价：

- T 是可逆的线性变换。
- 矩阵 A 可逆（A^{-1} 存在）。
- $\det(A) \neq 0$。
- A 是满秩矩阵：$\operatorname{rank}(A) = n$。
- A 的行向量组（或列向量组）线性无关。
- 齐次线性系统 $A\mathbf{x} = \mathbf{0}$ 只有平凡解 $\mathbf{x} = \mathbf{0}$。
- 对于任意 $\mathbf{b} \in \mathbb{R}^n$，非齐次系统 $A\mathbf{x} = \mathbf{b}$ 有唯一解。
- T 是单射（$\ker(T) = \{\mathbf{0}\}$）。
- T 是满射（$R(T) = \mathbb{R}^n$）。

8.6 Revisit Inverse, Determinant, Rank and Transpose

在学习了一种新的看待矩阵的视角后，我们不妨回头重新审视一些关于矩阵的基本概念，看看会有什么全新的理解。

逆序律 $(AB)^{-1} = B^{-1}A^{-1}$ 的几何解释

设 S 和 T 是可逆线性变换，矩阵分别为 A 和 B，则复合变换 $S \circ T$ 的矩阵为 AB。

- 要撤销“先做 T 再做 S”，逻辑上必须先撤销 S，再撤销 T。

- 代数上：$(S \circ T)^{-1} = T^{-1} \circ S^{-1}$。
- 矩阵上：$(AB)^{-1} = B^{-1}A^{-1}$。

这正是“穿脱原则”：操作顺序在求逆时需要反转。

行列式：变换的体积缩放因子

行列式 $\det(A)$ 测量了线性变换 $T : \mathbb{R}^n \to \mathbb{R}^n$ 对体积的缩放比例。

- 考虑 $\mathbb{R}^n$ 中以标准基向量为边的单位立方体。
- 施加变换 T（矩阵为 A）后，这个立方体被映射为一个平行多面体。
- 这个平行多面体的**有向体积**就是 $\det(A)$。

$\det(A)$ **值**	**几何意义**	**变换性质**
$\det(A) > 0$	保持定向的体积缩放	保持手性（如纯旋转、拉伸）
$\det(A) < 0$	改变定向的体积缩放	反转手性（如反射）
$\det(A) = 0$	体积坍缩为零	变换降维，不可逆
$\|\det(A)\| = 1$	体积不变	等距变换（旋转、反射）
$\|\det(A)\| > 1$	体积放大	整体拉伸
$\|\det(A)\| < 1$	体积缩小	整体压缩

行列式乘法定理 $\det(AB) = \det(A)\det(B)$ **的几何解释**

- 变换 T 将体积缩放 $\det(B)$ 倍。
- 接着变换 S 将体积再缩放 $\det(A)$ 倍。
- 复合变换 $S \circ T$ 的总缩放倍数是两次缩放的乘积：$\det(A)\det(B)$。

这解释了为什么行列式满足乘法定理：缩放因子的连续作用是相乘的。

矩阵的秩：值域的维度

从变换角度看，矩阵的秩 $\mathrm{rank}(A)$ 就是线性变换 T 的**值域的维度**。

- $\mathrm{rank}(A) = n$（满秩）：变换后仍然充满整个 $\mathbb{R}^n$，信息无损，可逆。
- $\mathrm{rank}(A) < n$：变换将空间压缩到一个更低维的子空间中，信息丢失，不可逆。
- **几何意义**：秩告诉你变换后的图像“占据了多少个维度”。例如，秩为 1 的变换将所有向量压缩到一条直线上；秩为 2 的变换将空间压缩到一个平面上。

矩阵的转置：对偶变换

若 $T : V \to W$ 的矩阵是 A，则其转置 A^T 对应的是**对偶变换** $T^* : W^* \to V^*$，其中 V^*, W^* 分别是 V, W 的对偶空间。

- **几何视角**：在 $\mathbb{R}^n$ 中配备标准内积时，A^T 可以看作“从行空间视角看同一个变换”。

- **重要性质**：$\text{rank}(A) = \text{rank}(A^T)$，意味着线性变换的值域维度等于其“行视角”下的有效方程数。

- **正交补关系**：$\text{RS}(A) = \text{CS}(A^T)$，$\text{CS}(A) = \text{RS}(A^T)$，这从变换角度解释了行空间和列空间的关系。

8.7 Rank-1 Matrix

理解矩阵的秩在线性变换层面的含义是非常重要的。一个 $m \times n$ 的矩阵代表一个从 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的变换，这个映射的定义域（输入域）是 $\mathbb{R}^n$，上域或者叫陪域是 $\mathbb{R}^m$，但值域（输出域）是 $\mathbb{R}^k$，k 是矩阵的秩。假设 $m \geq n$（或者更常见的，$m = n$），则值域的维度最高为 n，此时变换没有信息亏损。但若 $k < n$，则变换发生了信息亏损。

本节我们先来关注一类特殊的矩阵，秩正好为 1 的矩阵，我们叫**秩-1 矩阵**（rank-1 matrix）。这个名字我们并不陌生，事实上，早在介绍外积的时候，我们就抛下了两个结论：(1) **两个向量的外积得到的矩阵一定是秩-1 矩阵**。(2) 两个向量外积的迹等于它们的点积（当然此时两个向量的行数得相等）。理解第二个结论我们仍然需要等待，因为迹的概念将在下一章介绍矩阵的相似不变量的时候引出。但现在我们终于可以来讨论一下第一个结论。

事实上，不仅两个向量的外积得到的矩阵一定是秩-1 矩阵，**任意一个秩-1 矩阵都可以看作是两个向量的外积结果**。这是因为秩为 1 意味着这个矩阵的所有行（也等价于所有列）都是成比例的。值得注意的是，这种分解并**不是唯一的**。给定一个分解 $A = \mathbf{u}\mathbf{v}^T$，所有可能的分解是

$$A = (k\mathbf{u})\left(\frac{1}{k}\mathbf{v}^T\right), \quad k \neq 0$$

除此之外没有别的分解方式（在相差一个同时缩放的意义下唯一）。

理解了这个小结论后，我们希望从变换角度具体来看一下秩-1 矩阵带来的变换有什么特点，从而更深刻地理解秩的大小与变换的关系。理解了秩-1 矩阵，同样的分析可以应用于秩-2、秩-3、...、秩-k 矩阵。值得说明的是，由于秩相同的 $n \times n$ 矩阵和 $m \times n$ 矩阵在变换角度唯一的区别就是陪域的维度不同，但我们其实更加关心值域以及值域相较输入域发生的变化，因此在接下来的分析中我们都假设秩-1 矩阵为 $n \times n$ 的方阵。方阵的假设可以免去不重要的一般性，让我们的分析更加简便。当然，我们也会提及这个不重要的一般性如何在我们的分析中体现。

首先，按照我们已有的直觉，不满秩的方阵都意味着值域维度 k 小于定义域和陪域维度 n，图形在变换后发生维度坍缩，输入发生信息亏损。而秩-1 方阵是最极端的一种情况，值域的维度非常小，仅仅只有 1 维，这意味着值域的图形是一条过原点的直线。任何 $\mathbb{R}^n$ 中的向量左乘秩-1 矩阵得到的（依然是 $\mathbb{R}^n$ 中的）向量都在这条直线上。

下面从代数角度分析，设 $\mathbf{x}$ 是 $n \times 1$ 的向量，A 是 $n \times n$ 的矩阵，并且 $\text{rank}(A) = 1$，考虑 $A\mathbf{x}$ 这个向量的特点。我们说过，秩-1 矩阵可以写成两个向量的外积形式：

$$A = \mathbf{u}\mathbf{v}^T$$

其中 $\mathbf{u}$ 和 $\mathbf{v}$ 都是 $n \times 1$ 的非零列向量。写成元素形式：$A_{ij} = u_i v_j$。

下面我们计算 $A\mathbf{x}$，根据矩阵乘法的结合律：

$$A\mathbf{x} = (\mathbf{u}\mathbf{v}^T)\mathbf{x} = \mathbf{u}(\mathbf{v}^T\mathbf{x})$$

注意，$\mathbf{v}^T\mathbf{x}$ 是一个标量（向量的内积），记它为 $c = \mathbf{v}^T\mathbf{x} \in \mathbb{R}$。于是：

$$A\mathbf{x} = c\mathbf{u}$$

代数推导不仅验证了我们直觉层面的思考，还进一步具体地告诉我们那条过原点的直线为 $c\mathbf{u}$。

类似的，我们可以分析秩为 2 的矩阵结构并推广到 k。但这里我们需要用到一个结论，这个结论此处我们会不加说明地直接使用，在之后介绍矩阵分解的时候我们会再次遇到它并给予证明：**一个秩-k 矩阵可以被分解成 k 个秩-1 矩阵的和**。

一个 $n \times n$ 矩阵 A 秩为 2 意味着它可以写成两个秩-1 矩阵的和（但不唯一）：

$$A = \mathbf{u}_1\mathbf{v}_1^T + \mathbf{u}_2\mathbf{v}_2^T$$

其中 $\{\mathbf{u}_1, \mathbf{u}_2\}$ 线性无关，$\{\mathbf{v}_1, \mathbf{v}_2\}$ 线性无关（否则秩可能小于 2）。更一般地，可以表示为：

$$A = UV^T$$

其中 U 和 V 都是 $n \times 2$ 矩阵且列满秩（秩为 2）。

计算 $A\mathbf{x}$，用双外积形式：

$$A\mathbf{x} = (\mathbf{u}_1\mathbf{v}_1^T + \mathbf{u}_2\mathbf{v}_2^T)\mathbf{x} = \mathbf{u}_1(\mathbf{v}_1^T\mathbf{x}) + \mathbf{u}_2(\mathbf{v}_2^T\mathbf{x})$$

记标量：

$$c_1 = \mathbf{v}_1^T\mathbf{x}, \quad c_2 = \mathbf{v}_2^T\mathbf{x}$$

则：

$$A\mathbf{x} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2$$

秩-2 矩阵 A 把整个 $\mathbb{R}^n$ 映射到一个通过原点的二维平面。这是一个从 n 维到 2 维的"坍缩"，但比秩 1 的退化程度轻一些（保留了二维信息）。具体来说，输入 $\mathbf{x}$ 在 $\mathbf{v}_1, \mathbf{v}_2$ 方向上的分量（系数 c_1, c_2）被提取出来，然后按这些系数组合 $\mathbf{u}_1, \mathbf{u}_2$ 得到输出。这意味着 $\mathbf{x}$ 在垂直于 $\{\mathbf{v}_1, \mathbf{v}_2\}$ 所张成的空间的方向上的信息全部丢失（映射到零）。

同样的分析框架可以扩展到任意秩 k。对于秩-k 方阵 A：

$$A = \sum_{i=1}^{k} \mathbf{u}_i\mathbf{v}_i^T = UV^T$$

其中 U, V 是 $n \times k$ 矩阵且秩为 k。此时：

$$A\mathbf{x} = \sum_{i=1}^{k} (\mathbf{v}_i^T\mathbf{x})\mathbf{u}_i$$

结果向量一定落在 $\text{Span}\{\mathbf{u}_1, ..., \mathbf{u}_k\}$ 这个 k 维子空间中。

8.8 Low-rank Matrix

我们把秩**远小于** $\min(m, n)$ 的矩阵称为**低秩矩阵**（low-rank matrix）。低秩矩阵代表信息损失很大的线性变换。它的零空间很大，并且不可逆。尽管如此，很多时候我们却希望主动地压缩信息，例如在数据压缩、降噪、推荐系统等应用场景中。这时我们不需要每一个细节都完美保留的无损信息集，而是一个保留了主要特征，但有选择地丢弃了部分细节的信息集。这个时候我们的问题就变成了如何选择保留哪些信息丢弃哪些信息。或者说，如果我们一定要压缩 r 个维度的信息，我们要选择哪 r 个维度让它们成为零空间？或者更具体的，我们怎么判断哪些维度的信息是所谓"更主要的特征"？这些问题我们在下一章学习了特征值的概念后便可解答。

低秩矩阵相关的应用很多，例如低秩近似（Low-Rank Approximation）、矩阵补全（Matrix Completion）、鲁棒主成分分析（Robust PCA）、机器学习与深度学习中的低秩权重矩阵等。可见，虽然低秩矩阵代表了高度的信息亏损，但它却仍然因为独特的性质深受研究者的喜爱：

1. **压缩性**：存储和传输成本大幅下降。
2. **可解释性**：低秩对应“少数隐藏因子”生成数据，有物理或业务意义。
3. **计算效率**：低秩矩阵乘法、求逆、分解都更快。
4. **理论保证**：低秩假设使病态问题变得可解（如矩阵补全）。
5. **泛化能力**：在机器学习中，低秩结构可看作正则化，防止过拟合。

总结而言，低秩是现实世界中许多高维数据的内在简约性的数学体现——用很少的“基”就能近似表示复杂数据，从而带来计算和存储上的巨大优势。

例题 8.1 (低秩近似) 低秩近似的问题是这样的：给定一个大矩阵 A（例如 $m \times n, m, n$ 很大），想找一个秩 r 较小的矩阵 B 来近似 A，使得误差 $\|A - B\|$ 尽可能小。

8.9 Reflections, Rotations and Stretches in $\mathbb{R}^2$

线性变换可以很复杂，但任何复杂的线性变换都可以看作是由几种最基本、最纯粹的几何变换组合而成。因此，在熟悉和探索线性变换的过程中我们往往先从一些比较特殊的基础变换入手。

$\mathbb{R}^2$ 上的线性变换 T 可以用 2×2 矩阵 A 表示：

$$T(\mathbf{x}) = A\mathbf{x}, \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

其中 A 的列向量分别是 $T(\mathbf{e}_1)$ 与 $T(\mathbf{e}_2)$。

(1). 反射变换 (Reflections)

反射是将点关于过原点的直线作镜像映射。

1. 关于坐标轴的反射
   - 关于 x 轴反射：
   $$A_x = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$
   - 关于 y 轴反射：
   $$A_y = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$
   - 关于直线 $y = x$ 反射：
   $$A_{y=x} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$
2. 关于倾斜直线 $y = (\tan\theta)x$ 的反射

   设直线的单位方向向量为 $\mathbf{u} = (\cos\theta,\ \sin\theta)$，则反射矩阵为
   $$A_{\text{refl}}(\theta) = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}.$$
   也可以写作 $A = 2\mathbf{u}\mathbf{u}^T - I$。

反射矩阵均为正交矩阵（$A^TA = I$），且

$$\det A_{\text{refl}} = -1.$$

这符合几何直观：反射保持面积不变（面积缩放因子为 1），但改变了定向（行列式为负）。

(2). 旋转变换 (Rotations)

绕原点逆时针旋转 θ 角的变换矩阵为

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

这是一个正交矩阵：$R_\theta^{-1} = R_{-\theta} = R_\theta^T$。

$$\det R_\theta = \cos^2\theta + \sin^2\theta = 1.$$

这符合直觉：旋转既不改变面积，也不改变定向。

(3). 拉伸变换 (Stretches/Scaling)

沿坐标轴方向的伸缩变换。

1. 均匀拉伸（缩放）
   缩放因子为 $k > 0$：
   $$S_k = \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix} = kI.$$
2. 非均匀拉伸
   沿 x 方向缩放 a 倍，沿 y 方向缩放 b 倍：
   $$S_{a,b} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}.$$

$$\det S_k = k^2, \quad \det S_{a,b} = ab.$$

这正是面积缩放的比例因子：均匀拉伸时面积变为 k^2 倍；非均匀拉伸时面积变为 ab 倍。当 $a, b > 0$ 时，行列式为正，定向不变。

8.10 Orthogonal Matrices: Rigid Transformations

我们已经研究了 $\mathbb{R}^2$ 中的三种基本几何变换，可这种试图遍历所有方阵代表的基本变换类型的研究过程效率还是太低了，我们现在进一步缩小研究重点，来研究一下一类特殊的方阵所代表的线性变换——正交矩阵。

一个 $n \times n$ 实矩阵 P 称为**正交矩阵 (orthogonal matrix)**，如果满足

$$P^TP = I \quad \text{或等价地} \quad P^T = P^{-1}.$$

特别地，正交矩阵一定是可逆矩阵。

若 A 是正交矩阵，则 $|\det A| = 1$。更确切地说，

$$\det A = \pm 1.$$

因为：
$$1 = \det I = \det(P^TP) = (\det P)^2 \quad \Rightarrow \quad \det P = \pm 1.$$

矩阵的正交性与向量间的正交性有紧密联系：

一个 $n \times n$ 矩阵 P 是正交矩阵当且仅当它的列向量两两正交且每个列向量的长度为 1；也就是说，当且仅当 P 的列构成 $\mathbb{R}^n$ 的一组标准正交基。

证明思路：记 $P = [\mathbf{p}_1\ \mathbf{p}_2\ \cdots\ \mathbf{p}_n]$，则
$$P^TP = \begin{pmatrix} \mathbf{p}_1^T \\ \mathbf{p}_2^T \\ \vdots \\ \mathbf{p}_n^T \end{pmatrix} [\mathbf{p}_1\ \mathbf{p}_2\ \cdots\ \mathbf{p}_n] = (\mathbf{p}_i \cdot \mathbf{p}_j)_{n\times n}.$$
因此 $P^TP = I$ 等价于 $\mathbf{p}_i \cdot \mathbf{p}_j = \delta_{ij}$。

对于一个实 $n \times n$ 矩阵 A，下列陈述等价：

1. A 是正交矩阵；
2. 对任意 $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$，$(A\mathbf{u}) \cdot (A\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$；
3. 对任意 $\mathbf{u} \in \mathbb{R}^n$，$\|A\mathbf{u}\| = \|\mathbf{u}\|$（保持向量长度）。

证明 8.1

(1) ⇒ (2)：$(A\mathbf{u}) \cdot (A\mathbf{v}) = (A\mathbf{u})^T(A\mathbf{v}) = \mathbf{u}^TA^TA\mathbf{v} = \mathbf{u}^T\mathbf{v} = \mathbf{u} \cdot \mathbf{v}$。

(2) ⇒ (3)：取 $\mathbf{v} = \mathbf{u}$，则 $\|A\mathbf{u}\|^2 = (A\mathbf{u}) \cdot (A\mathbf{u}) = \mathbf{u} \cdot \mathbf{u} = \|\mathbf{u}\|^2$。

(3) ⇒ (1)：利用极化恒等式 $\mathbf{u}\cdot\mathbf{v} = \frac{1}{4}(\|\mathbf{u}+\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{v}\|^2)$，结合条件可推出 $(A\mathbf{u})\cdot(A\mathbf{v}) = \mathbf{u}\cdot\mathbf{v}$，从而 $\mathbf{u}^TA^TA\mathbf{v} = \mathbf{u}^T\mathbf{v}$ 对所有 $\mathbf{u}, \mathbf{v}$ 成立，故 $A^TA = I$。

一个从 $\mathbb{R}^n$ 到 $\mathbb{R}^n$ 的线性变换保持 $\mathbb{R}^n$ 的几何结构（即长度和角度）**当且仅当**它由一个正交矩阵表示。换句话说，正交矩阵对应的线性变换是 $\mathbb{R}^n$ 中的**刚体运动**：

- 若 $\det P = 1$，则为**旋转**变换（保持定向）；
- 若 $\det P = -1$，则为**包含反射**的变换（反转定向）。

在 $\mathbb{R}^2$ 中，一个正交矩阵所代表的变换只有可能是**旋转**或者**反射**。设 A, B 为 $\mathbb{R}^2$ 中的正交矩阵，则：

- 两个旋转的复合仍是旋转：$\det(AB) = 1 \cdot 1 = 1$。
- 两个反射的复合是旋转：$\det(AB) = (-1) \cdot (-1) = 1$。
- 一个旋转与一个反射的复合是反射：$\det(AB) = 1 \cdot (-1) = -1$。

代数验证：$|AB| = |A||B|$。

几何解释：连续两次反射相当于一次旋转；反射会改变定向，而旋转不会。

8.11 Reflections and Rotations in $\mathbb{R}^3$

在 $\mathbb{R}^3$ 中，正交矩阵代表的线性变换的分类更加丰富，因为定向反转不一定对应纯反射，还可能是旋转与反射的组合。

旋转变换 (Rotation)

绕 z 轴旋转 θ 角（保持定向）：

$$R_z(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \det = 1.$$

一般地，绕任意过原点的轴旋转的矩阵均满足 $\det = 1$。

反射变换 (Reflection)

关于 xy 平面的反射（反转定向）：

$$F_{xy} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad \det = -1.$$

关于任意过原点的平面的反射矩阵均有 $\det = -1$。

旋转-反射复合变换 (Rotation–reflection)

这是三维空间特有的类型：先绕某轴旋转，再关于垂直于该轴的平面反射（顺序可交换）。例如：

$$B = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad \det = -1.$$

该变换相当于绕 z 轴旋转 θ 角后再关于 xy 平面反射。

8.12 Exercises

9 Linear Transformations II 线性变换 II（从标准基推广至任意基）

9.1 The Matrix $A_T^{B\to C}$

在本节中，我们为从一个有限维向量空间 V 到另一个有限维向量空间 W 的线性变换 $T: V \to W$ 寻找一个矩阵表示。空间 V 和 W 可能不是欧几里得空间。只要为 T 的定义域 V 引入一组基 B，并为 T 的到达域 W 引入一组基 C，那么元素 $v \in V$ 和 $T(v) \in W$ 就可以分别用关于 B 和 C 的坐标向量 $(v)_B$ 和 $(T(v))_C$ 来表示。由此得到的关于基 B 和 C 表示 $T: V \to W$ 的矩阵记为 $A_T^{B\to C}$。

我们首先从以下定理开始：

定理 9.1 设 V 是一个有限维向量空间，T 是从 V 到向量空间 W 的一个线性变换。那么 T 完全由它在一组基上的作用方式所决定。

证明 9.1 设 $\dim(V) = n$，并令 $B = \{v_1, v_2, \ldots, v_n\}$ 是 V 的一组基。那么任何 $v \in V$ 都可以唯一地表示为这些基向量的线性组合：$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$。因此，由 T 的线性性，

$$T(v) = T(\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n) = \alpha_1 T(v_1) + \alpha_2 T(v_2) + \cdots + \alpha_n T(v_n).$$

也就是说，如果 $v \in V$ 被表示为基向量的线性组合，那么像 $T(v)$ 就是基向量之像的同一线性组合。因此，如果我们知道 T 如何作用于基向量，我们就知道 T 如何作用于所有 $v \in V$。 □

在 V 和 W 都是有限维向量空间的特定情况下，并且只要为 V 引入了基 B、为 W 引入了基 C，这个结果就允许我们找到线性变换 T 的一个矩阵表示。

更具体地说，设 $\dim(V) = n$，$\dim(W) = m$，相应的基由 $B = \{v_1, v_2, \ldots, v_n\}$ 和 $C = \{w_1, w_2, \ldots, w_m\}$ 给出。此外，令 $(v)_B$ 是 $v \in V$ 关于 B 基的坐标向量，令 $(T(v))_C$ 是 $T(v)$ 关于 C 基的坐标向量。那么，通过处理这些坐标向量（而不是向量本身），我们可以找到一个矩阵，使得

$$(T(\mathbf{v}))_C = \mathbf{A}_T^{B\to C}(\mathbf{v})_B.$$

下面的定理告诉我们如何找到这个矩阵。

定理 9.2 设 $T: V \to W$ 是一个从 n 维向量空间 V 到 m 维向量空间 W 的线性变换。令 $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ 表示定义域 V 的一组基，$C = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ 表示到达域 W 的一组基。进一步，令 $\mathbf{A}_T^{B\to C}$ 是那个 $m \times n$ 矩阵，其列是 B-基向量的像关于 C 基的坐标

$$(T(\mathbf{v}_1))_C, (T(\mathbf{v}_2))_C, \ldots, (T(\mathbf{v}_n))_C;$$

也就是说，

$$\mathbf{A}_T^{B\to C} = \begin{pmatrix} (T(\mathbf{v}_1))_C & (T(\mathbf{v}_2))_C & \cdots & (T(\mathbf{v}_n))_C \end{pmatrix}.$$

那么，对于每个 $\mathbf{v} \in V$，都有

$$(T(\mathbf{v}))_C = \mathbf{A}_T^{B\to C}(\mathbf{v})_B.$$

注意，如果 V 是 $\mathbb{R}^n$，W 是 $\mathbb{R}^m$，并且 B 和 C 分别是 V 和 W 的标准基，那么 $\mathbf{A}_T^{B\to C}$ 就变成了上一章引入的矩阵 $\mathbf{A}_T$；也就是说，

$$\mathbf{A}_T^{B\to C} = \begin{pmatrix} (T(\mathbf{v}_1))_C & (T(\mathbf{v}_2))_C & \cdots & (T(\mathbf{v}_n))_C \end{pmatrix} = \begin{pmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \end{pmatrix} = \mathbf{A}_T.$$

定理 9.3 设 $T : V \to W$ 是有限维向量空间之间的一个线性变换，并设 B 和 C 分别是 T 的定义域 V 和到达域 W 的基。那么

$$\ker(T) = N(A_T^{B\to C}), \quad R(T) = CS(A_T^{B\to C}),$$

其中，$N(A_T^{B\to C})$ 中每个向量的坐标是相对于 V 的 B 基而言的，而 $CS(A_T^{B\to C})$ 中每个向量的坐标是相对于 W 的 C 基而言的。

这里，$N(A_T^{B\to C})$ 表示矩阵 $A_T^{B\to C}$ 的零空间（核），$CS(A_T^{B\to C})$ 表示其列空间。

9.2 Change of Basis and Transition Matrix

考虑欧几里得空间 $\mathbb{R}^n$。假设向量 $v_1, v_2, \ldots, v_n$ 构成 $\mathbb{R}^n$ 的一组基 B。那么，正如我们所见，任何 $x \in \mathbb{R}^n$ 都可以唯一地写成基 B 中向量的线性组合：

$$x = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n.$$

向量

$$(x)_B = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}_B$$

就是 x 关于基 $B = \{v_1, v_2, \ldots, v_n\}$ 的坐标向量。注意，只要明确坐标 $\alpha_1, \alpha_2, \ldots, \alpha_n$ 是相对于基 B 的，下标 B 可以从上述等式的右侧省略。换句话说，我们也可以写成

$$(x)_B = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

在 B 是 $\mathbb{R}^n$ 的标准基 $\{e_1, e_2, \ldots, e_n\}$ 的特殊情况下，坐标向量 $(x)_B$ 与 x 本身重合。这是因为如果 $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$，那么

$$x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n, \quad \text{因此} \quad (x) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

在实际操作中，为了找到给定向量 x 关于基 $B = \{v_1, v_2, \ldots, v_n\}$ 的坐标，我们只需要求解线性方程组

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = x.$$

这个方程组可以写成如下形式

$$\begin{pmatrix} v_1 & v_2 & \dots & v_n \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = x,$$

其中 $\begin{pmatrix} v_1 & v_2 & \dots & v_n \end{pmatrix}$ 是以基 B 中的向量为列的矩阵。将这个矩阵记为 P_B，即

$$P_B = \begin{pmatrix} v_1 & v_2 & \dots & v_n \end{pmatrix},$$

并利用

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = (x)_B \quad 和 \quad x = (x),$$

上述方程变为

$$P_B(x)_B = (x).$$

矩阵 P_B 将 x 关于 B-基的坐标 $(x)_B$ 与 x 关于标准基的坐标 (x) 联系起来。正因如此，P_B 被称为**从 B-坐标到标准坐标的过渡矩阵**。注意，$n \times n$ 矩阵 P_B 是可逆的，因为它的列构成 $\mathbb{R}^n$ 的一组基，这意味着 P_B 的秩是 n。所以我们也可以写成

$$(x)_B = P_B^{-1}(x).$$

那么矩阵 P_B^{-1} 就是**从标准坐标到 B-坐标的过渡矩阵**。

9.3 The Transition Matrix $P_{B\to B'}$

更一般地，假设我们已知 $\mathbb{R}^n$ 的一个基 B，另一个基 B'，以及向量 $\mathbf{v} \in \mathbb{R}^n$ 的坐标 $(\mathbf{v})_B$。那么，从 B-坐标到 B'-坐标的**过渡矩阵** $\mathbf{P}_{B\to B'}$，以及向量 $\mathbf{v}$ 的坐标 $(\mathbf{v})_{B'}$，可以按如下方式计算：

首先，利用 $(\mathbf{v}) = \mathbf{P}_B(\mathbf{v})_B$ 将坐标从 B-坐标转换为标准坐标，然后利用 $(\mathbf{v})_{B'} = \mathbf{P}_{B'}^{-1}(\mathbf{v})$ 将坐标从标准坐标转换为 B'-坐标。对初始坐标向量 $(\mathbf{v})_B$ 的综合效果是：

$$(\mathbf{v})_{B'} = \mathbf{P}_{B'}^{-1}\mathbf{P}_B(\mathbf{v})_B,$$

这意味着从 B-坐标到 B'-坐标的**过渡矩阵**为：

$$\mathbf{P}_{B\to B'} = \mathbf{P}_{B'}^{-1}\mathbf{P}_B.$$

另一种计算 $\mathbf{P}_{B\to B'}$ 的方法由以下定理提供。

定理 9.4 设 B 和 B' 是 $\mathbb{R}^n$ 的两个基，其中第一个基是：

$$B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}.$$

那么，从 B-坐标到 B'-坐标的过渡矩阵 $\mathbf{P}_{B\to B'}$ 由下式给出：

$$\mathbf{P}_{B\to B'} = \begin{pmatrix} (\mathbf{v}_1)_{B'} & (\mathbf{v}_2)_{B'} & \dots & (\mathbf{v}_n)_{B'} \end{pmatrix},$$

其中矩阵 $\mathbf{P}_{B\to B'}$ 的列由 B-基向量关于 B' 基的坐标构成。

注意，在 B' 是 $\mathbb{R}^n$ 的标准基的特殊情况下，推导 $\mathbf{P}_{B\to B'}$ 的两种方法都得到预期的、从 B-坐标到标准坐标的过渡矩阵 $\mathbf{P}_B$。对于前一种方法，我们有：

$$\mathbf{P}_{B\to B'} = \mathbf{P}_{B'}^{-1}\mathbf{P}_B = \mathbf{I}^{-1}\mathbf{P}_B = \mathbf{I}\mathbf{P}_B = \mathbf{P}_B.$$

还需要注意，计算 $P_{B\to B'}$ 的第二种方法直接适用于任何有限维向量空间 V，因为其中可能没有“标准”基的概念。相比之下，第一种方法——即使用 $P_{B\to B'} = P_{B'}^{-1}P_B$ ——预设了 V 存在一个标准基，因此除非我们指定 V 的一个基来充当“标准”基，否则该方法不适用。下面的例子说明了这一点。

例题 9.1 考虑集合：

$$V = \{f : \mathbb{R} \to \mathbb{R} \mid f(x) = a + bx，对于某些 a, b \in \mathbb{R}\},$$

它在函数的逐点加法和数乘标准运算下构成一个向量空间。以下集合 B, C 和 D 都是 V 的基：

$$\begin{aligned}
B &= \{f_1, f_2\} \quad 其中 f_1(x) = 1,\ f_2(x) = x,\\
C &= \{g_1, g_2\} \quad 其中 g_1(x) = 2,\ g_2(x) = 1 + x,\\
D &= \{h_1, h_2\} \quad 其中 h_1(x) = 2 + x,\ h_2(x) = 1 + 2x.
\end{aligned}$$

现在考虑一个元素 $f \in V$，其关于 C 基的坐标向量为 $(f)_C = \begin{pmatrix} 3 \\ 5 \end{pmatrix}_C$。求 $(f)_D$。

解法一：直接从定义出发

首先，我们可以不使用任何过渡矩阵，直接从基本原理出发解决这个问题：陈述 $(f)_C = \begin{pmatrix} 3 \\ 5 \end{pmatrix}_C$ 意味着 $f = 3g_1 + 5g_2$，即对于所有 $x \in \mathbb{R}$：

$$f(x) = 3g_1(x) + 5g_2(x) = 3(2) + 5(1 + x) = 11 + 5x.$$

为了找到 f 的坐标 $(f)_D$，我们只需要将 f 表示为 D-基向量的线性组合。换句话说，我们需要找到标量 α_1, α_2，使得对于所有 $x \in \mathbb{R}$：

$$11 + 5x = \alpha_1(2 + x) + \alpha_2(1 + 2x) = (2\alpha_1 + \alpha_2) + x(\alpha_1 + 2\alpha_2).$$

由于这个等式必须对所有 $x \in \mathbb{R}$ 都成立，它必须关于 x 恒等，这意味着：

$$\begin{cases} 2\alpha_1 + \alpha_2 = 11 \\ \alpha_1 + 2\alpha_2 = 5. \end{cases}$$

解这个联立方程组，我们得到 $\alpha_1 = \frac{17}{3}$ 和 $\alpha_2 = -\frac{1}{3}$，从而：

$$(f)_D = \begin{pmatrix} \frac{17}{3} \\ -\frac{1}{3} \end{pmatrix}_D.$$

这种方法的问题在于它不是系统性的。如果我们被问到另一个 V 中的元素，就需要重新开始。

解法二：使用系统方法（过渡矩阵）

系统化的方法在于通过使用从 C-坐标到 D-坐标的过渡矩阵 $P_{C\to D}$ 来获得 $(f)_D$。当然，我们意识到 V 没有“标准”基，除非我们指定 B, C 或 D 中的一个来扮演这个角色。因此，让我们从计算 $P_{C\to D}$ 的第二种方法开始，因为这种方法不预设 V 存在一个标准基。

利用结果 $P_{C\to D}=\begin{pmatrix}(g_1)_D & (g_2)_D\end{pmatrix}$，我们需要做的就是把 C-基向量 g_1,g_2 表示为 D-基向量 h_1,h_2 的线性组合。

从 g_1 开始，令 $g_1=a_1h_1+a_2h_2$。那么，对于所有 $x\in\mathbb{R}$，我们需要满足：

$$g_1(x)=a_1h_1(x)+a_2h_2(x),\quad \text{即}2=a_1(2+x)+a_2(1+2x).$$

由于这个等式必须关于 x 恒等，我们得到：

$$\begin{cases}2=2a_1+a_2\\0=a_1+2a_2,\end{cases}$$

其解为 $a_1=\frac{4}{3},a_2=-\frac{2}{3}$。因此 $(g_1)_D=\begin{pmatrix}\frac{4}{3}\\-\frac{2}{3}\end{pmatrix}_D$。

类似地，对于 g_2，令 $g_2=b_1h_1+b_2h_2$。那么，对于所有 $x\in\mathbb{R}$，我们需要满足：

$$g_2(x)=b_1h_1(x)+b_2h_2(x),\quad \text{即}1+x=b_1(2+x)+b_2(1+2x).$$

通过类似的论证，我们得到联立方程组：

$$\begin{cases}1=2b_1+b_2\\1=b_1+2b_2,\end{cases}$$

其解为 $b_1=\frac{1}{3},b_2=\frac{1}{3}$。因此 $(g_2)_D=\begin{pmatrix}\frac{1}{3}\\\frac{1}{3}\end{pmatrix}_D$。

于是：

$$\mathbf{P}_{C\to D}=\begin{pmatrix}\frac{4}{3}&\frac{1}{3}\\-\frac{2}{3}&\frac{1}{3}\end{pmatrix},$$

并且：

$$(f)_D=\mathbf{P}_{C\to D}(f)_C=\begin{pmatrix}\frac{4}{3}&\frac{1}{3}\\-\frac{2}{3}&\frac{1}{3}\end{pmatrix}\begin{pmatrix}3\\5\end{pmatrix}_C=\begin{pmatrix}\frac{17}{3}\\-\frac{1}{3}\end{pmatrix}_D.$$

这样我们就得到了与之前相同的答案。

解法三：通过指定标准基（第一种方法）

为了完整性，让我们用第一种方法计算 $\mathbf{P}_{C\to D}$。如前所述，这预设了我们指定，例如，B 为 V 的“标准”基。然后我们需要找到从 C-坐标到“标准”坐标的过渡矩阵 $\mathbf{P}_C$，以及从 D-坐标到“标准”坐标的过渡矩阵 $\mathbf{P}_D$，最后应用公式 $\mathbf{P}_{C\to D}=\mathbf{P}_D^{-1}\mathbf{P}_C$。

指定 B 为“标准”基，并将一切用 B-坐标表示，通过观察我们有：

$$B=\left\{\begin{pmatrix}1\\0\end{pmatrix}_B,\begin{pmatrix}0\\1\end{pmatrix}_B\right\},$$
$$C=\left\{\begin{pmatrix}2\\0\end{pmatrix}_B,\begin{pmatrix}1\\1\end{pmatrix}_B\right\},$$
$$D=\left\{\begin{pmatrix}2\\1\end{pmatrix}_B,\begin{pmatrix}1\\2\end{pmatrix}_B\right\}.$$

既然我们已经决定将 B 视为“标准基”，甚至可以去掉下标 B：

$$B=\left\{\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix}\right\},$$

$$C = \left\{ \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\},$$

$$D = \left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}.$$

与欧几里得空间完全一样，上述表达式意味着：

$$\mathbf{P}_C = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{和} \quad \mathbf{P}_D = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

因此：

$$\mathbf{P}_D^{-1} = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix},$$

并且：

$$P_{C\to D} = P_D^{-1} P_C = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{4}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} \end{pmatrix}.$$

结论 $(f)_D = P_{C\to D}(f)_C = \begin{pmatrix} \frac{4}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 3 \\ 5 \end{pmatrix}_C = \begin{pmatrix} \frac{17}{3} \\ -\frac{1}{3} \end{pmatrix}_D$ 再次得到验证。

9.4 Change of Basis and Linear Transformations

给定一个线性变换 $T : \mathbb{R}^n \to \mathbb{R}^m$，我们已经知道存在一个对应的矩阵 A_T，即

$$A_T = \Big(T(\mathbf{e}_1) \quad T(\mathbf{e}_2) \quad \ldots \quad T(\mathbf{e}_n) \Big),$$

使得对于所有 $\mathbf{x} \in \mathbb{R}^n$，有 $T(\mathbf{x}) = A_T\mathbf{x}$。

更一般地，给定 T 的定义域 $\mathbb{R}^n$ 的一组基

$$B = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$$

以及 T 的到达域 $\mathbb{R}^m$ 的一组基

$$C = \{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_m\},$$

我们也知道存在一个矩阵 $A_T^{B\to C}$，即

$$A_T^{B\to C} = \Big((T(\mathbf{f}_1))_C \quad (T(\mathbf{f}_2))_C \quad \cdots \quad (T(\mathbf{f}_n))_C \Big),$$

使得对于所有 $\mathbf{x} \in \mathbb{R}^n$，有 $(T(\mathbf{x}))_C = A_T^{B\to C}(\mathbf{x})_B$。

正如预期的那样，矩阵 $A_T^{B\to C}$ 和 A_T 之间存在着一种关系，这种关系涉及到实现 $\mathbb{R}^n$ 和 $\mathbb{R}^m$ 中相应坐标变化的过渡矩阵 P_B 和 P_C。

从对所有 $\mathbf{x} \in \mathbb{R}^n$ 都成立的事实出发：

$$(T(\mathbf{x})) = A_T(\mathbf{x}),$$

并利用关系式 $(T(\mathbf{x})) = P_C(T(\mathbf{x}))_C$ 和 $(\mathbf{x}) = P_B(\mathbf{x})_B$，我们得到对于所有 $\mathbf{x} \in \mathbb{R}^n$：

$$P_C(T(\mathbf{x}))_C = A_T P_B(\mathbf{x})_B.$$

在等式左边乘以 P_C^{-1}，得到：

$$(T(\mathbf{x}))_C = P_C^{-1} A_T P_B(\mathbf{x})_B.$$

因此，由于我们还有 $(T(\mathbf{x}))_C = A_T^{B\to C}(\mathbf{x})_B$，我们得到矩阵 A_T 和 $A_T^{B\to C}$ 之间的如下关系：

$$A_T^{B\to C} = P_C^{-1} A_T P_B.$$

这个公式 $A_T^{B\to C} = P_C^{-1} A_T P_B$ 揭示了线性变换在不同基下的矩阵表示是如何转换的：

- P_B：将 $\mathbf{x}$ 从 B-坐标转换到标准坐标。
- A_T：在标准坐标下执行变换 T。
- P_C^{-1}：将结果从标准坐标转换到 C-坐标。

9.5 Similar Matrix

特别令人感兴趣的一种特殊情况是：线性变换的定义域和到达域是同一个欧几里得空间，即 $T : \mathbb{R}^n \to \mathbb{R}^n$，并且定义域和到达域的基重合，即 $B = C$。在这种情况下，连接 A_T 和 $A_T^{B\to C}$ 的一般方程简化为：

$$A_T^{B\to B} = P_B^{-1} A_T P_B.$$

此时，矩阵 A_T 和 $A_T^{B\to B}$ 被称为**相似**矩阵。

一般地，一个方阵 N 被称为与另一个方阵 M **相似**，如果存在一个可逆方阵 P，使得

$$N = P^{-1} M P.$$

注意，我们也有 $PNP^{-1} = M$，这意味着存在一个可逆矩阵 Q（即 $Q = P^{-1}$），使得

$$M = Q^{-1} N Q.$$

因此，M 也与 N 相似。相似的矩阵（如 A_T 和 $A_T^{B\to B}$）代表了同一个线性变换 T 在不同基下的表示。

更一般地，对于 $\mathbb{R}^n$ 的任意基 B 和 C，我们有：

$$A_T^{B\to B} = P_B^{-1} A_T P_B \quad \text{和} \quad A_T^{C\to C} = P_C^{-1} A_T P_C.$$

从第一个方程解出 A_T 并将得到的表达式代入第二个方程，我们得到：

$$A_T^{C\to C} = P_C^{-1}(P_B A_T^{B\to B} P_B^{-1}) P_C = (P_C^{-1} P_B) A_T^{B\to B} (P_B^{-1} P_C).$$

现在，回顾之前的内容：

$$P_{C\to B} = P_B^{-1} P_C \quad \text{和} \quad P_{C\to B}^{-1} = P_{B\to C} = P_C^{-1} P_B.$$

因此，上述关系变为：

$$A_T^{C\to C} = P_{C\to B}^{-1} A_T^{B\to B} P_{C\to B},$$

这确立了矩阵 $A_T^{B\to B}$ 和 $A_T^{C\to C}$ 是相似的事实。第一个矩阵代表了 T 关于基 B 的表示，而第二个矩阵代表了 T 关于基 C 的表示。

- 相似矩阵描述了**同一个线性变换**在**不同基**下的坐标表示。
- 相似变换 $N = P^{-1} M P$ 本质上是一个坐标变换：先用 P 变到新基，应用变换 M，再用 P^{-1} 变回原坐标系（或反之）。
- 相似矩阵具有许多共同的不变量，如**行列式**、**迹**、**特征多项式**（因而**特征值**）等，这反映了变换本身的固有属性与基的选择无关。

相似矩阵的概念强调了变换的某些性质是**独立于坐标系选择**的，这是我们接下来要研究的重点！

9.6 Diagonalisable Linear Transformations

我们已经知道，同一个线性变换在不同基底下对应不同的矩阵表示。假设一个线性变换在某个特定的基底下，其变换矩阵是对角矩阵（当然这样的基不一定存在），那么在这个基底下观察该线性变换就很容易理解它的效果。

给定一个线性变换 $T : \mathbb{R}^n \to \mathbb{R}^n$，假设我们能够找到 $\mathbb{R}^n$ 的一组基 B，使得表示 T 的矩阵 $A_T^{B\to B}$ 是对角矩阵，即

$$A_T^{B\to B} = \begin{pmatrix} k_1 & 0 & \cdots & 0 \\ 0 & k_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & k_n \end{pmatrix},$$

其中 $k_1, k_2, \ldots, k_n$ 是给定的常数。在特定的基 $B = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$ 下，很容易理解变换 T 的效果：T 将每个基向量 $\mathbf{f}_i$ 拉伸了因子 k_i 倍。

定义 9.1 (矩阵的可对角化) 一个 $n \times n$ 矩阵 A 被称为**可对角化的**，如果它与某个对角矩阵相似。也就是说，存在一个可逆矩阵 P，使得

$$P^{-1}AP = D = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n),$$

其中 D 是一个对角矩阵。

等价地，我们可以写为

$$A = PDP^{-1}.$$

对角化意味着我们可以找到一个基底 B（由 P 的列向量给出），使得线性变换 $T(\mathbf{x}) = A\mathbf{x}$ 在该基底下的作用非常简单：它仅仅沿着每个基底方向进行伸缩（伸缩系数即为对角矩阵 D 对角线上的特征值）。这个基底中的向量就是 A 的特征向量。

对角化将矩阵 A 分解为三个矩阵的乘积：$A = PDP^{-1}$。这种分解具有重要的代数优势：

- **矩阵幂的计算**：$A^k = PD^kP^{-1}$，而 D^k 很容易计算（只需将对角元分别取 k 次幂）。
- **矩阵函数的计算**：对于多项式函数 f，$f(A) = Pf(D)P^{-1}$，其中 $f(D)$ 是对角矩阵 $f(\mathrm{diag}(\lambda_1, \ldots, \lambda_n)) = \mathrm{diag}(f(\lambda_1), \ldots, f(\lambda_n))$。
- **解微分方程组**：对角化可将线性常系数微分方程组解耦。

9.7 Eigenvalues, Eigenvectors and Eigenspaces

在研究线性变换 $T : \mathbb{R}^n \to \mathbb{R}^n$ 时，一个自然的问题是：是否存在一些特殊的向量，这些向量在变换 T 的作用下保持方向不变，仅被拉伸或压缩？

这种特殊的向量对于理解线性变换的几何本质至关重要。如果能够找到这样一组向量作为基底，那么线性变换在该基底下将具有非常简单的形式（对角矩阵），从而大大简化分析。

设 A 是一个 $n \times n$ 矩阵，对应的线性变换为 $T(\mathbf{x}) = A\mathbf{x}$。

定义 9.2 (特征值与特征向量) 对于一个非零向量 $\mathbf{v} \in \mathbb{R}^n$ 和一个标量 $\lambda \in \mathbb{R}$（或 $\mathbb{C}$），如果满足

$$A\mathbf{v} = \lambda\mathbf{v},$$

则称 λ 是矩阵 A 的一个**特征值**，$\mathbf{v}$ 是相应于特征值 λ 的一个**特征向量**。

几何解释：特征向量 $\mathbf{v}$ 在变换 A 的作用下方向保持不变（或反向，若 $\lambda < 0$），仅被拉伸或压缩了 $|\lambda|$ 倍。

从定义 $A\mathbf{v} = \lambda\mathbf{v}$ 可得：

$$A\mathbf{v} - \lambda\mathbf{v} = \mathbf{0} \quad \Rightarrow \quad (A - \lambda I)\mathbf{v} = \mathbf{0},$$

其中 I 是 $n \times n$ 单位矩阵。

为了使这个齐次线性方程组有非零解 $\mathbf{v} \neq \mathbf{0}$，矩阵 $(A - \lambda I)$ 必须是奇异的，即：

$$\det(A - \lambda I) = 0.$$

这个方程称为矩阵 A 的**特征方程 (eigenvalue equation)**。

将 $\det(A - \lambda I)$ 展开，得到一个关于 λ 的 n 次多项式：

$$p(\lambda) = \det(A - \lambda I) = (-1)^n \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0.$$

这个多项式称为 A 的**特征多项式 (characteristic polynomial equation)**。

特征值就是特征多项式的根。根据代数基本定理，一个 $n \times n$ 矩阵在复数域中恰好有 n 个特征值（计入重数）。

对于一个特定的特征值 λ，所有满足 $(A - \lambda I)\mathbf{v} = \mathbf{0}$ 的向量 $\mathbf{v}$ 构成一个子空间。

定义 9.3 (特征空间) 矩阵 A 相应于特征值 λ 的**特征空间**定义为：

$$E_\lambda = \{\mathbf{v} \in \mathbb{R}^n \mid (A - \lambda I)\mathbf{v} = \mathbf{0}\} = \ker(A - \lambda I).$$

特征空间 E_λ 就是齐次线性方程组 $(A - \lambda I)\mathbf{x} = \mathbf{0}$ 的解空间。

例题 9.2 给定 $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$：

1. **特征方程**：

$$\det(A - \lambda I) = \det\begin{pmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{pmatrix} = (3-\lambda)^2 - 1 = \lambda^2 - 6\lambda + 8 = 0.$$

2. **特征值**：解得 $\lambda_1 = 2, \lambda_2 = 4$。

3. **特征向量与特征空间**：

   - 对于 $\lambda_1 = 2$：解 $(A - 2I)\mathbf{v} = \mathbf{0}$：

   $$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Rightarrow \quad \mathbf{v} = t\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad t \neq 0.$$

   特征空间 $E_2 = \text{span}\left\{\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right\}$。

   - 对于 $\lambda_2 = 4$：解 $(A - 4I)\mathbf{v} = \mathbf{0}$：

   $$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Rightarrow \quad \mathbf{v} = t\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad t \neq 0.$$

   特征空间 $E_4 = \text{span}\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}$。

9.8 Relationship between Diagonalisation and Eigenvectors

矩阵的**特征向量**指的是变换后只被拉伸的方向，而**对角化**指的是针对同一个线性变换重新建系（也就是更换基），使得该变换作用于新的基全部为最简单的拉伸效果，可以理解为为同一个线性变换找一组代数上最便捷的基。可以看出这两个概念的定义并无直接依赖关系，但我们也知道它们之间有很紧密的联系：如果一个 $n \times n$ 方阵可对角化，那么说明我们能找到 n 个线性独立的向量方向在变换后只被拉伸，因此可以就以这 n 个向量为基重新建系。所以如果一个 $n \times n$ 方阵可对角化，那么等价于存在 n 个线性独立的特征向量。本节旨在进一步梳理两者的关系，也为下一节讨论对角化的条件做铺垫。

定理 9.5 (对角化的特征向量判据) *设 A 为 $n \times n$ 矩阵，则以下陈述等价：*

1. *A 可对角化（即存在可逆矩阵 P 和对角矩阵 D 使得 $A = PDP^{-1}$）。*
2. *A 有 n 个线性无关的特征向量。*
3. *存在由 A 的特征向量构成的一组基。*

这是一个**充要条件**，它建立了对角化与线性无关特征向量个数的直接联系。

注意，线性无关的特征向量最多只可能有 n 个，但是特征向量方向可以远多于 n 个（甚至无限多）。只要能从这些方向中选出 n 个线性无关的向量，矩阵就可对角化。想象三维空间中一个过零点的平面上的所有向量均在变换后只产生拉伸效果，此时特征向量的方向有无数个，但由于共面最多只能挑出 2 个线性无关的特征向量，因此无法对角化。再举个极端的例子，考虑三维单位矩阵，显然每个非零向量都是特征向量，并且可以找到 3 个线性无关的特征向量建系，因此可以对角化。这不仅说明了当矩阵可以对角化时特征向量的方向个数有可能大于 n，也说明了当矩阵不可以对角化时不代表不存在特征向量，甚至特征向量的方向个数也有可能大于 n。因此，准确来说，可否对角化与**线性无关的特征向量的个数**有关。

9.9 Conditions of Diagonalisation

本节我们讨论一个方阵何时能对角化。上节我们说过一个 $n \times n$ 的矩阵 A 可对角化等价于它有 n 个线性无关的特征向量。这个定理将问题转化为：何时 A 有 n 个线性无关的特征向量？我们首先看一个充分条件：

定理 9.6 *属于不同特征值的特征向量是线性无关的。*

这个定理的直接推论是：如果 $n \times n$ 矩阵 A **有 n 个互异的实特征值** $\lambda_1, \ldots, \lambda_n$，则对应的特征向量 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 线性无关，从而构成 $\mathbb{R}^n$ 的一组基 B。此时：

- 变换矩阵 $P_B = (\mathbf{v}_1 \; \mathbf{v}_2 \; \cdots \; \mathbf{v}_n)$ 可逆
- 在基 B 下，T 的表示矩阵为对角矩阵 $A_T^{B \to B} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$
- 有关系式 $A_T = P_B A_T^{B \to B} P_B^{-1}$，即 A 可对角化

这是最朴素的情况，下面我们思考存在相同特征值的情况。拿三维举例，假设特征值 a 的重数是 2。有两个特征向量方向被拉伸的倍数一样无所谓，重点在于它们是不是独立的。因为按之前的算法一定能解出 n 个特征值，对应 n 组特征向量。互异的特征值不用管它们，因为对应的特征向量一定是相互线性独立的。我们只需要担心相同的特征值，也就是有重数的特征值。我们关心这 2 个相同的特征值对应的特征向量是否也线性独立。如果它们线性相关，那说明它们本质是同一组

特征向量（同一个方向）。同样的思想拓展到 n 维，我们关心那 k 个相同的特征值对应的特征向量共同构成的空间的维度，如果是 k，那么说明确实是有 k 个线性独立的特征向量方向恰好拉伸倍数是相同的。如果小于 k，说明有些特征向量对应的特征值相同不是巧合，而是因为它们是线性相关的。我们可以定义所谓"**代数重数 (Algebraic Multiplicity)**"和"**几何重数 (Geometric Multiplicity)**"的概念来阐述这一段逻辑：

定义 9.4 (代数重数与几何重数) 设 λ_0 是 $n \times n$ 矩阵 A 的一个实特征值。

- λ_0 的**代数重数**是指最大的整数 k，使得 $(\lambda - \lambda_0)^k$ 是特征多项式 $|A - \lambda I|$ 的因子。
- λ_0 的**几何重数**是指对应特征空间 $N(A - \lambda_0 I)$ 的维数。

重要性质：

- 对于任何实特征值 λ_0，其几何重数至少为 1（因为 $|A - \lambda_0 I| = 0$ 意味着 $A - \lambda_0 I$ 不是满列秩）。
- 几何重数不超过代数重数：几何重数 $\leq$ 代数重数。
- 如果 A 有 n 个互异的实特征值，则每个特征值的代数重数和几何重数都等于 1。

结合代数基本定理（n 次多项式恰有 n 个复根，计入重数），我们得到：

定理 9.7 一个矩阵 A 可对角化当且仅当满足以下两个条件：

1. 特征多项式方程只产生实特征值（即所有特征值都是实数）。

2. 每个特征值的几何重数等于其代数重数。

注意我们这里的表述是"几何重数等于其代数重数"，而不是"代数重数等于其几何重数"或"两个重数相等"，是因为在判断顺序上我们始终是以特征值为基准。特征值的代数重数是先求得的，作为已知条件再引导我们去求代数重数大于 1 的特征值的几何重数，以此来与代数重数比对。

1. **特征值全为实数**：对角化要求存在实数域上的特征向量基。如果特征值为复数，则对应的特征向量也是复的，不能在实数域 $\mathbb{R}^n$ 中形成基。
2. **重数相等**：这是关键。对于每个特征值 λ，代数重数 k 表示它在特征多项式中作为根的重数，几何重数 g 表示对应线性无关特征向量的个数。只有当 $g = k$ 时，我们才能从该特征值对应的特征空间中找到 k 个线性无关的特征向量。如果对每个特征值都满足这一条件，则所有特征空间提供的线性无关特征向量总数等于 n，从而构成一组基。

考虑三种 2×2 矩阵的情形：

1. **无实特征值**：特征多项式无实根，不可对角化（在实数域上）。
2. **有重特征值但几何重数 $<$ 代数重数**：例如 $\lambda = 1$（代数重数 2），但特征空间维数为 1。此时只有 1 个线性无关的特征向量，总数 < 2，不可对角化。
3. **有重特征值但几何重数 $=$ 代数重数**：例如 $\lambda = 1$（代数重数 2），特征空间维数也为 2。此时有 2 个线性无关的特征向量，可对角化。

总结而言，对角化的本质是寻找一组由特征向量构成的基。实现这一目标的障碍来自两方面：

- **特征值非实**：在实数域上无法获得足够的实特征向量。
- **几何重数不足**：即使特征值是实数且重数较高，但如果对应的特征空间维数太小，也无法提供足够的线性无关特征向量。

9.10 Orthogonal Diagonalisation

一个矩阵 A 被称为**正交可对角化的 (orthogonally diagonalisable)**，如果存在一个正交矩阵 P，使得

$$P^{-1}AP = P^TAP = D,$$

其中 D 是一个对角矩阵。

注意：

- A 可对角化意味着 P 的列向量构成 $\mathbb{R}^n$ 的一组由 A 的特征向量组成的基。
- 正交可对角化的额外要求是 P 是正交矩阵，即其列向量构成 $\mathbb{R}^n$ 的一组**标准正交基**，且每个列向量都是 A 的特征向量。

由此可得以下定理：

定理 9.8 矩阵 A 正交可对角化当且仅当存在 $\mathbb{R}^n$ 的一组由 A 的特征向量组成的标准正交基。

定理 9.9 矩阵 A 正交可对角化当且仅当 A 是对称矩阵（即 $A^T = A$）。

这个定理是实对称矩阵谱定理的核心。它意味着：

- 任何实对称矩阵都可以通过正交变换对角化。
- 反之，如果一个矩阵可以正交对角化，那么它一定是对称的。

由于一个矩阵可对角化要求其特征值全是实数，上述定理立即推出以下推论：

推论 9.1 如果 A 是对称矩阵，那么它的所有特征值都是实数。

此外，上述定理还意味着：即使对称矩阵 A 有重特征值，对应于不同特征值的特征空间也是正交的。否则，A 不可能有一组由特征向量组成的标准正交基，而这正是正交对角化所必需的。由此可进一步推出：

定理 9.10 如果矩阵 A 是对称的，那么对应于不同特征值的特征向量是正交的。

证明 9.2 设 $\mathbf{v}_1$ 和 $\mathbf{v}_2$ 是 A 的特征向量，对应特征值 $\lambda_1 \neq \lambda_2$。考虑内积：

$$\lambda_1(\mathbf{v}_1 \cdot \mathbf{v}_2) = (A\mathbf{v}_1) \cdot \mathbf{v}_2 = \mathbf{v}_1^T A^T \mathbf{v}_2 = \mathbf{v}_1^T A \mathbf{v}_2 = \mathbf{v}_1 \cdot (A\mathbf{v}_2) = \lambda_2(\mathbf{v}_1 \cdot \mathbf{v}_2).$$

由于 $\lambda_1 \neq \lambda_2$，必有 $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$，即 $\mathbf{v}_1 \perp \mathbf{v}_2$。

对于一个对称矩阵 A，构造特征向量标准正交基 B 的方法如下：

1. 对 A 的每个特征值 λ，求出对应的特征空间 $N(A - \lambda I)$。
2. 在每个特征空间内部应用格拉姆-施密特正交化过程：
   - 如果特征空间是一维的，只需确保其基向量为单位长度。
   - 如果特征空间是多维的，则需要完整的格拉姆-施密特过程来产生一组标准正交基。
3. 将所有特征空间的标准正交基合并在一起。由于不同特征值对应的特征空间是正交的，且每个特征空间内部已经正交化，因此得到的向量集合构成 $\mathbb{R}^n$ 的一组标准正交基。

9.11 Symmetric Matrices and Quadratic Forms

二次型 (quadratic form) 是多个变量的二次齐次多项式。它在优化、物理和几何中都有重要应用。

两个变量 x 和 y 的二次型形式为：

$$q(x,y) = ax^2 + 2cxy + by^2.$$

这可以写成矩阵形式：

$$q(x,y) = \mathbf{x}^T\mathbf{A}\mathbf{x},$$

其中 $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$，$\mathbf{A}$ 是对称矩阵：

$$\mathbf{A} = \begin{pmatrix} a & c \\ c & b \end{pmatrix}.$$

验证展开矩阵乘积 $\mathbf{x}^T\mathbf{A}\mathbf{x}$ 确实得到 $ax^2 + 2cxy + by^2$。

类似地，n 个变量 $x_1, x_2, \ldots, x_n$ 的二次型形式为：

$$q(x_1, x_2, \ldots, x_n) = \mathbf{x}^T\mathbf{A}\mathbf{x},$$

其中 $\mathbf{A}$ 是 $n \times n$ 对称矩阵，$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$。

例题 9.3 *以下是三个变量的二次型：*

$$q(x_1, x_2, x_3) = 5x_1^2 + 10x_2^2 + 2x_3^2 + 4x_1x_2 + 2x_1x_3 - 6x_2x_3.$$

我们有 $q(x_1, x_2, x_3) = \mathbf{x}^T\mathbf{A}\mathbf{x}$*，其中* $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$*，*$\mathbf{A}$ *是对称矩阵：*

$$\mathbf{A} = \begin{pmatrix} 5 & 2 & 1 \\ 2 & 10 & -3 \\ 1 & -3 & 2 \end{pmatrix}.$$

从二次型表达式 $q(x_1, \ldots, x_n)$ 推导对称矩阵 $\mathbf{A}$ 的规则如下（反之亦然）：

- 对角线元素 a_{ii} 是二次项 x_i^2 的系数。
- 非对角线元素 $a_{ij} = a_{ji}$ 是交叉项 x_ix_j 系数的一半。

因此，无需进行矩阵乘法即可在二次型与其对称矩阵表示之间转换。

假设我们找到了正交矩阵 P 使对称矩阵 A 正交对角化，即 $P^T = P^{-1}$ 且 $P^TAP = D$，其中 D 是对角矩阵。

进行变量代换 $\mathbf{x} = P\mathbf{z}$，其中 P 被视为从特征向量标准正交基 B 的坐标到标准坐标的过渡矩阵 P_B，$\mathbf{z} = (\mathbf{x})_B$ 是 $\mathbf{x}$ 关于 B 基的坐标向量。则：

$$q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x} = (P\mathbf{z})^TA(P\mathbf{z}) = \mathbf{z}^T(P^TAP)\mathbf{z} = \mathbf{z}^TD\mathbf{z}.$$

设 D 的对角元素（即 A 的特征值）为 $\lambda_1, \lambda_2, \ldots, \lambda_n$（允许重复）。令 $\mathbf{z} = (\mathbf{x})_B = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$，则：

$$q(\mathbf{x}) = \lambda_1 X_1^2 + \lambda_2 X_2^2 + \cdots + \lambda_n X_n^2.$$

这就是二次型的**标准形**——仅包含平方项的线性组合。

我们关心二次型 $q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$（其中 $A^T = A$）可能取的所有值的集合。为此引入以下术语：

定义 9.5 (定号性) 设 $q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$（$A^T = A$）是二次型。

- $q(\mathbf{x})$ 是**正定的**，如果对所有 $\mathbf{x} \neq \mathbf{0}$ 有 $q(\mathbf{x}) > 0$，且 $q(\mathbf{0}) = 0$。
- $q(\mathbf{x})$ 是**半正定的**，如果对所有 $\mathbf{x}$ 有 $q(\mathbf{x}) \geq 0$。
- $q(\mathbf{x})$ 是**负定的**，如果对所有 $\mathbf{x} \neq \mathbf{0}$ 有 $q(\mathbf{x}) < 0$，且 $q(\mathbf{0}) = 0$。
- $q(\mathbf{x})$ 是**半负定的**，如果对所有 $\mathbf{x}$ 有 $q(\mathbf{x}) \leq 0$。
- $q(\mathbf{x})$ 是**不定的**，如果它既不是半正定也不是半负定；即存在 $\mathbf{x}_1, \mathbf{x}_2$ 使得 $q(\mathbf{x}_1) < 0$ 且 $q(\mathbf{x}_2) > 0$。

定理 9.11 设 $q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$，其中 $A^T = A$，则：

- $q(\mathbf{x})$ 正定 $\iff$ A 的所有特征值均为正。
- $q(\mathbf{x})$ 半正定 $\iff$ A 的所有特征值均为非负。
- $q(\mathbf{x})$ 负定 $\iff$ A 的所有特征值均为负。
- $q(\mathbf{x})$ 半负定 $\iff$ A 的所有特征值均为非正。
- $q(\mathbf{x})$ 不定 $\iff$ A 至少有一个正特征值和一个负特征值。
- ($\Rightarrow$) 若 $q(\mathbf{x})$ 正定，取 A 的对应于特征值 λ_i 的单位特征向量 $\mathbf{u}_i$，则：

  $$q(\mathbf{u}_i) = \mathbf{u}_i^T A \mathbf{u}_i = \mathbf{u}_i^T \lambda_i \mathbf{u}_i = \lambda_i \|\mathbf{u}_i\|^2 = \lambda_i > 0.$$

  故每个特征值 $\lambda_i > 0$。
- ($\Leftarrow$) 若所有 $\lambda_i > 0$，则 $q(\mathbf{x}) = \sum_{i=1}^n \lambda_i X_i^2 \geq 0$，且等号成立仅当所有 $X_i = 0$，即 $\mathbf{z} = \mathbf{0}$，从而 $\mathbf{x} = P\mathbf{z} = \mathbf{0}$。

其他情况的证明类似。

上述术语也直接延伸到对称矩阵 A 本身。因此，对称矩阵 A 是：

- **正定的 (positive definite)** 当且仅当其所有特征值为正。
- **半正定的 (positive semi-definite)** 当且仅当其所有特征值为非负。
- **负定的 (negative definite)** 当且仅当其所有特征值为负。
- **半负定的 (negative semi-definite)** 当且仅当其所有特征值为非正。
- **不定的 (indefinite)** 当且仅当至少有一个正特征值和一个负特征值。

关于 $\mathbb{R}^2$ 中的二次型，它们与**圆锥曲线 (conic sections)** 直接相关。回顾圆锥曲线是 $\mathbb{R}^2$ 中由形如

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

的笛卡尔方程描述的曲线，其中 A, B, C, D, E, F 是给定实数。

在这些曲线中，任何笛卡尔方程可写成更简单形式

$$ax^2 + 2cxy + by^2 = k$$

的圆锥曲线，都可以表达为涉及二次型的矩阵方程：

$$\mathbf{x}^T A\mathbf{x} = k,$$

其中

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad A = \begin{pmatrix} a & c \\ c & b \end{pmatrix} = A^T.$$

在此背景下，正交对角化技术可用于找到 $\mathbb{R}^2$ 的一组标准正交基 B，使得圆锥曲线相对于 B 坐标轴处于标准位置和方向。

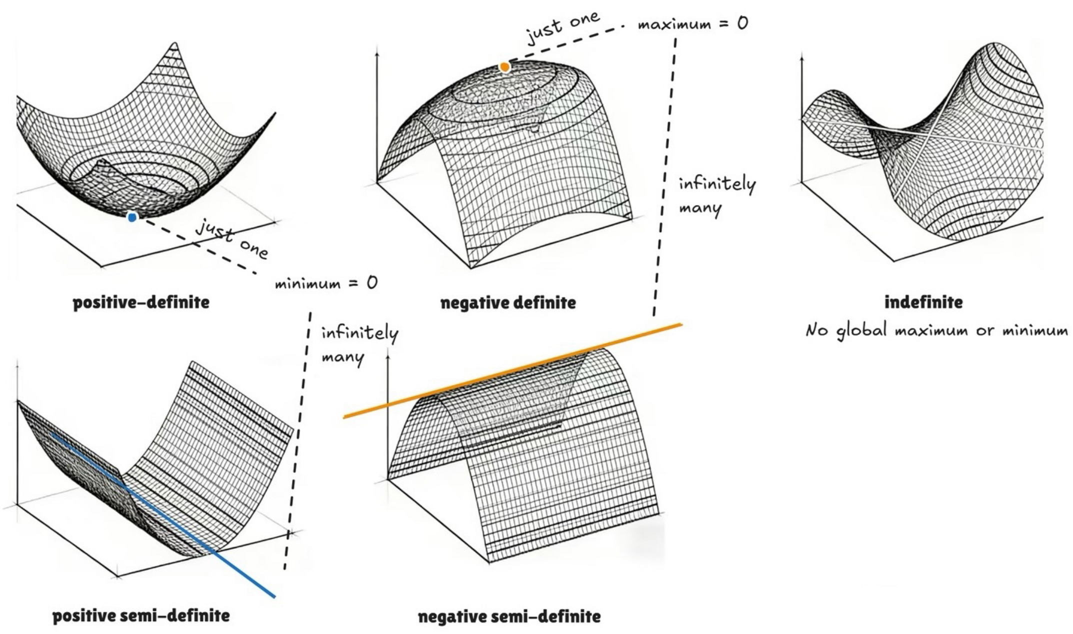

图 9: 对称矩阵和二次型

9.12 Trace and other Similarity Invariants

在线性代数中，我们关心哪些矩阵性质在相似变换下保持不变。这些性质称为**相似不变量**，它们描述了线性变换的本质特征，而不依赖于特定基的选择。**迹 (trace)**、行列式（determinant）和特征多项式是最重要的相似不变量。

设 A 是 $n \times n$ 矩阵，其迹定义为对角线元素之和：

$$\text{tr}(A) = \sum_{i=1}^{n} a_{ii}$$

以下是矩阵的迹的一些基本性质：

1. **线性性**：$\operatorname{tr}(cA + dB) = c\operatorname{tr}(A) + d\operatorname{tr}(B)$，其中 $c, d \in \mathbb{R}$。
2. **转置不变**：$\operatorname{tr}(A^T) = \operatorname{tr}(A)$。
3. **循环置换性**：对于可乘的矩阵 A, B，有

$$\operatorname{tr}(AB) = \operatorname{tr}(BA)$$

   更一般地，$\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB)$（只要乘积有意义）。

4. **相似不变性**：若 A 与 B 相似（即存在可逆 P 使得 $B = P^{-1}AP$），则

$$\operatorname{tr}(B) = \operatorname{tr}(A)$$

定理 9.12 (迹-特征值关系) 设 A 是 $n \times n$ 矩阵，特征值为 $\lambda_1, \lambda_2, \ldots, \lambda_n$（计入代数重数），则

$$\operatorname{tr}(A) = \sum_{i=1}^{n} \lambda_i$$

证明 9.3 考虑特征多项式：

$$p(\lambda) = \det(\lambda I - A) = \lambda^n - \operatorname{tr}(A)\lambda^{n-1} + \cdots + (-1)^n \det(A)$$

另一方面，$p(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n)$。比较 λ^{n-1} 项的系数即得结论。

迹可以理解为线性变换 $T(\mathbf{x}) = A\mathbf{x}$ 的"平均伸缩率"或"总膨胀系数"的度量。在微分方程中，迹与解的散度有关。

行列式也是一个重要的相似不变量。若 $B = P^{-1}AP$，则

$$\det(B) = \det(P^{-1}AP) = \det(P^{-1})\det(A)\det(P) = \det(A)$$

因为 $\det(P^{-1}) = 1/\det(P)$。

定理 9.13 (行列式-特征值关系)

$$\det(A) = \prod_{i=1}^{n} \lambda_i$$

其中 λ_i 是 A 的特征值（计入代数重数）。

$|\det(A)|$ 表示线性变换 $T(\mathbf{x}) = A\mathbf{x}$ 对体积的缩放因子。$\det(A)$ 的符号表示定向是否改变：- $\det(A) > 0$：保持定向 - $\det(A) < 0$：反转定向 - $\det(A) = 0$：变换降维（不可逆）

特征多项式也是一个重要的相似不变量。若 $B = P^{-1}AP$，则

$$\det(\lambda I - B) = \det(\lambda I - P^{-1}AP) = \det(P^{-1}(\lambda I - A)P) = \det(\lambda I - A)$$

所以相似矩阵有相同的特征多项式。

设 A 的特征多项式为：

$$p(\lambda) = \lambda^n - c_1\lambda^{n-1} + c_2\lambda^{n-2} - \cdots + (-1)^n c_n$$

则：

- $c_1 = \operatorname{tr}(A)$（迹）

- $c_n = \det(A)$（行列式）
- c_k 是所有 k 阶主子式之和（在 $k=2$ 时为 $\sum_{i<j}\lambda_i\lambda_j$）

其他相似不变量包括：

- **秩**
若 $B = P^{-1}AP$，则 $\operatorname{rank}(B) = \operatorname{rank}(A)$。因为可逆矩阵相乘不改变秩。
- **特征值及其代数重数**
相似矩阵有完全相同的特征值（包括重数）。
- **特征空间维数（几何重数）**
对于每个特征值 λ，$\dim\ker(A-\lambda I)$ 在相似变换下不变。
- **最小多项式**
矩阵 A 的最小多项式 $m_A(\lambda)$ 是满足 $m_A(A)=0$ 的最低次首一多项式。相似矩阵有相同的最小多项式。

定理 9.14 (Cayley-Hamilton 定理) 每个矩阵都满足其自身的特征方程。即若 $p(\lambda) = \det(\lambda I - A)$，则 $p(A) = 0$。

9.13 Exercises

10 Linear Transformations III 线性变换 III（关注一类特殊的变换：投影）

在实向量空间中，投影是一种非常重要且实用的线性变换。本章将系统介绍投影的各个方面，从最简单的情形开始，逐步深入到一般子空间上的投影，并探讨投影矩阵的性质。这些内容将为后续学习谱分解和奇异值分解 (SVD) 奠定基础。

10.1 Projections

线性变换的种类有很多，但我们说过我们试图找到一些基本变换使得再复杂的变换也不过是若干基本变换的复合。事实上，之后我们会了解到（参考奇异值分解），任何一个线性变换都可以分解成如下操作的叠加：旋转——变维拉伸——旋转。这也同时意味着任何一个矩阵都可以分解成三个矩阵的乘积，它们分别对应了这些基本变换。作为基本变换，旋转和拉伸我们并不陌生（当然这里的“拉伸”也包括反射，即负的拉伸），但维度改变的变换（主要指值域维度小于输入域维度的**降维**变换，因为陪域维度大于输入域维度的所谓“升维”变换没有太大意义）我们却还未相对系统地讨论过任何具体的变换类别。本章我们就来重点关注一类特殊且重要的变换：**投影（projection）**。可以说，它是最重要的一类降维变换。

在几何中，**投影**是将一个向量映射到另一个向量或子空间上的操作。最简单的投影形式是将向量 $\mathbf{v}$ 投影到向量 $\mathbf{u}$ 上，这可以看作在 $\mathbf{u}$ 方向上的“阴影”。

设 $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$，且 $\mathbf{u} \neq \mathbf{0}$，则向量 $\mathbf{v}$ 在向量 $\mathbf{u}$ 方向上的**标量投影**为：

$$\mathrm{comp}_{\mathbf{u}}(\mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|}$$

其中 $\mathbf{u} \cdot \mathbf{v}$ 是标准内积，$\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}}$ 是 $\mathbf{u}$ 的长度。

相应的**向量投影**（或称为 $\mathbf{v}$ 在 $\mathbf{u}$ 上的正交投影）为：

$$\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \left(\frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}}\right)\mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|^2}\mathbf{u}$$

投影具有以下基本性质：

1. **线性性**：投影操作是线性的，即对于任意标量 c_1, c_2 和向量 $\mathbf{v}_1, \mathbf{v}_2$，有

$$\mathrm{proj}_{\mathbf{u}}(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1\mathrm{proj}_{\mathbf{u}}(\mathbf{v}_1) + c_2\mathrm{proj}_{\mathbf{u}}(\mathbf{v}_2)$$

2. **幂等性**：投影的投影等于自身，即

$$\mathrm{proj}_{\mathbf{u}}(\mathrm{proj}_{\mathbf{u}}(\mathbf{v})) = \mathrm{proj}_{\mathbf{u}}(\mathbf{v})$$

10.2 Projections onto Subspaces

设 U 和 W 是实向量空间 V 的两个子空间，满足 $V = U \oplus W$（直和分解）。即，对于任意向量 $\mathbf{v} \in V$，存在唯一的分解：

$$\mathbf{v} = \mathbf{u} + \mathbf{w}, \quad \mathbf{u} \in U,\ \mathbf{w} \in W$$

从 V 到 U 沿 W 方向的**平行投影**是线性变换 $T: V \to V$，定义为：

$$T(\mathbf{v}) = \mathbf{u}$$

其中 $\mathbf{u}$ 是 $\mathbf{v}$ 在直和分解 $V = U \oplus W$ 中的 U 分量。

- 该变换将整个空间 V 映射到子空间 U 上。
- 沿 W 方向：对于任意 $\mathbf{w} \in W$，有 $T(\mathbf{w}) = \mathbf{0}$，即 W 中的向量被投影到零向量。

设 T 是从 V 到 U 沿 W 方向的投影变换，则：

1. **值域**：$\mathrm{Im}(T) = U$
2. **核空间**：$\mathrm{Ker}(T) = W$
3. **直和分解**：$V = \mathrm{Im}(T) \oplus \mathrm{Ker}(T)$

当 $V = \mathbb{R}^n$ 时，投影变换 T 可以用矩阵 $P \in \mathbb{R}^{n\times n}$ 表示，使得 $T(\mathbf{v}) = P\mathbf{v}$。
关键性质：投影矩阵 P 满足**幂等性（idempotence）**，即：

$$P^2 = P$$

这样的矩阵称为**幂等矩阵（idempotent matrix）**。

从几何角度理解：对一个向量 $\mathbf{v}$ 进行一次投影得到 $P\mathbf{v}$，再次应用同样的投影，结果应保持不变：

$$P(P\mathbf{v}) = P\mathbf{v} \quad \text{对所有 } \mathbf{v} \in \mathbb{R}^n$$

这等价于矩阵条件 $P^2 = P$。

由此可进一步推出，对于任意正整数 n，有：

$$P^n = P$$

10.3 Idempotent Matrix

定义 10.1 (幂等矩阵) *实方阵 P 称为**幂等矩阵**，如果满足 $P^2 = P$。*

定理 10.1 (投影与幂等矩阵的等价性) *设 P 是 $n \times n$ 实矩阵，则以下条件等价：*

1. *P 表示一个投影变换（即存在子空间 $U, W \subseteq \mathbb{R}^n$ 使得 $\mathbb{R}^n = U \oplus W$ 且 P 是从 $\mathbb{R}^n$ 到 U 沿 W 方向的投影）*
2. *P 是幂等矩阵（$P^2 = P$）*
3. *$\mathbb{R}^n = R(P) \oplus N(P)$，其中 $R(P)$ 表示 P 的值域，$N(P)$ 表示 P 的零空间*

当这些条件成立时，P 投影到 $R(P)$ 上，且沿 $N(P)$ 方向。

特殊幂等矩阵：

- 零矩阵 O：表示到零子空间的投影
- 单位矩阵 I：表示到整个空间 $\mathbb{R}^n$ 的投影（实际上是恒等变换）

定理 10.2 *幂等矩阵的特征值只能是 0 或 1。*

证明 10.1 *设 $P\mathbf{v} = \lambda\mathbf{v}$，其中 $\mathbf{v} \neq \mathbf{0}$。由于 $P^2 = P$，有：*

$$P^2\mathbf{v} = P\mathbf{v} \Rightarrow P(\lambda\mathbf{v}) = \lambda\mathbf{v} \Rightarrow \lambda^2\mathbf{v} = \lambda\mathbf{v}$$

因此 $\lambda^2 = \lambda$，解得 $\lambda = 0$ 或 $\lambda = 1$。

- 特征值 1 对应的特征向量属于 $R(P)$（值域）
- 特征值 0 对应的特征向量属于 $N(P)$（零空间）

定理 10.3 幂等矩阵可对角化。

证明 10.2 由于 $\mathbb{R}^n = R(P) \oplus N(P)$，可选择 $R(P)$ 的一组基和 $N(P)$ 的一组基，它们的并构成 $\mathbb{R}^n$ 的一组基。在这组基下，P 的矩阵表示为对角矩阵，对角线元素为 1（对应 $R(P)$ 的基向量）和 0（对应 $N(P)$ 的基向量）。

定理 10.4 对于幂等矩阵 P，有 $\mathrm{tr}(P) = \mathrm{rank}(P)$。

证明 10.3 由于 P 可对角化且特征值为 0 或 1，迹等于特征值之和，即特征值 1 的个数。而特征值 1 的几何重数等于 $R(P)$ 的维数，即 $\mathrm{rank}(P)$。

定理 10.5 幂等矩阵的行列式只能是 0 或 1。

证明 10.4 行列式等于特征值的乘积。由于特征值只能是 0 或 1，行列式只能是 0（当至少有一个特征值为 0）或 1（当所有特征值均为 1，此时 $P = I$）。

定理 10.6 可逆的幂等矩阵必为单位矩阵。

证明 10.5 若 P 可逆且幂等，则 $P^2 = P$ 两边左乘 P^{-1} 得 $P = I$。

这意味着除了单位矩阵外，所有幂等矩阵都是奇异的（不可逆）。

对于 2×2 实矩阵，所有幂等矩阵具有以下形式之一：

1. 零矩阵：$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$
2. 单位矩阵：$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
3. 一般形式：$\begin{pmatrix} a & b \\ c & 1-a \end{pmatrix}$，其中 $a, b, c \in \mathbb{R}$ 满足 $a^2 + bc = a$

推导：设 $P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ 满足 $P^2 = P$，则：

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 = \begin{pmatrix} a^2 + bc & ab + bd \\ ac + cd & bc + d^2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

比较对应元素可得：

$$a^2 + bc = a \tag{10.1}$$
$$ab + bd = b \tag{10.2}$$
$$ac + cd = c \tag{10.3}$$
$$bc + d^2 = d \tag{10.4}$$

从方程 (10.2) $ab+bd=b$:

$$ab+bd-b=0 \;\Rightarrow\; b(a+d-1)=0$$

由此得:

$$b=0 \quad \text{或} \quad a+d=1$$

从方程 (10.3) $ac+cd=c$:

$$ac+cd-c=0 \;\Rightarrow\; c(a+d-1)=0$$

由此得:

$$c=0 \quad \text{或} \quad a+d=1$$

情况 1:$a+d=1$(即 $d=1-a$)
将 $d=1-a$ 代入方程 (10.1) 和 (10.4):
方程 (10.1) 保持不变:$a^2+bc=a$。
方程 (10.4):$bc+(1-a)^2=1-a$,展开得:

$$bc+1-2a+a^2=1-a$$

$$bc+a^2-2a=-a$$
$$bc+a^2=a$$

这与方程 (10.1) 完全相同。

因此,当 $d=1-a$ 时,四个方程等价于单个条件:

$$a^2+bc=a$$

矩阵形式为:

$$P=\begin{pmatrix} a & b \\ c & 1-a \end{pmatrix}, \quad a^2+bc=a$$

情况 2:$a+d\neq 1$
根据 (10.2) 和 (10.3),此时必须有 $b=0$ 且 $c=0$。
代入 $b=0, c=0$:
方程 (10.1) 变为:$a^2=a\Rightarrow a(a-1)=0\Rightarrow a=0$ 或 $a=1$
方程 (10.4) 变为:$d^2=d\Rightarrow d(d-1)=0\Rightarrow d=0$ 或 $d=1$
但需满足 $a+d\neq 1$,排除以下组合:

- $(a,d)=(0,1)$:$0+1=1$,违反 $a+d\neq 1$ 假设
- $(a,d)=(1,0)$:$1+0=1$,违反 $a+d\neq 1$ 假设

剩余有效组合:

- $(a,d)=(0,0)$:满足 $0+0=0\neq 1$,对应矩阵 $P=\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$
- $(a,d)=(1,1)$:满足 $1+1=2\neq 1$,对应矩阵 $P=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

例题 10.1 (幂等矩阵的例子) 以下矩阵是幂等矩阵：

$$A = \begin{pmatrix} 4 & -2 \\ 6 & -3 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & -3 & -5 \\ -1 & 4 & 5 \\ 1 & -3 & -4 \end{pmatrix}$$

可直接验证 $A^2 = A$，$B^2 = B$。

10.4 Parallel Projections

设 $\mathbb{R}^n = U \oplus W$（直和分解），其中 U, W 均为 $\mathbb{R}^n$ 的子空间。平行投影 $P_{U,W} : \mathbb{R}^n \to U$ 将向量沿 W 方向投影到 U 上。

方法一：通过基变换（通用方法）

设 $\dim(U) = m$，$\dim(W) = n - m$。选择 $\mathbb{R}^n$ 的一组基 $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{v}_{m+1}, \ldots, \mathbf{v}_n\}$，其中：

- $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ 是 U 的一组基
- $\{\mathbf{v}_{m+1}, \ldots, \mathbf{v}_n\}$ 是 W 的一组基

在这组基 $\mathcal{B}$ 下，平行投影变换的表示矩阵非常简单。对于任意向量 $\mathbf{x} \in \mathbb{R}^n$，设其在基 $\mathcal{B}$ 下的坐标为：

$$[\mathbf{x}]_{\mathcal{B}} = (c_1, \ldots, c_m, c_{m+1}, \ldots, c_n)^T$$

其中 $\sum_{i=1}^{m} c_i \mathbf{v}_i \in U$，$\sum_{i=m+1}^{n} c_i \mathbf{v}_i \in W$。

投影操作保留 U 分量，将 W 分量置零：

$$[P_{U,W}(\mathbf{x})]_{\mathcal{B}} = (c_1, \ldots, c_m, 0, \ldots, 0)^T$$

因此，在基 $\mathcal{B}$ 下，投影变换的矩阵表示为：

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix}_{n \times n}$$

其中 I_m 是 $m \times m$ 单位矩阵。

设 $M_{\mathcal{B}} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n] \in \mathbb{R}^{n \times n}$ 是以 $\mathcal{B}$ 中向量为列构成的过渡矩阵。则标准基下的投影矩阵为：

$$P = M_{\mathcal{B}} [T]_{\mathcal{B}}^{\mathcal{B}} M_{\mathcal{B}}^{-1}$$

将 $M_{\mathcal{B}}$ 分块为 $M_{\mathcal{B}} = [A \mid B]$，其中：

- $A = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_m] \in \mathbb{R}^{n \times m}$ 是 U 的基矩阵
- $B = [\mathbf{v}_{m+1} \ \cdots \ \mathbf{v}_n] \in \mathbb{R}^{n \times (n-m)}$ 是 W 的基矩阵

则：

$$P = [A \mid B] \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} A_{(1)}^{-1} \\ A_{(2)}^{-1} \end{pmatrix}$$

其中 $\begin{pmatrix} A_{(1)}^{-1} \\ A_{(2)}^{-1} \end{pmatrix}$ 是 $M_{\mathcal{B}}^{-1}$ 的分块形式，$A_{(1)}^{-1} \in \mathbb{R}^{m \times n}$ 为前 m 行，$A_{(2)}^{-1} \in \mathbb{R}^{(n-m) \times n}$ 为后 $n - m$ 行。

计算得：

$$P = [A \mid B] \begin{pmatrix} I_m \\ 0 \end{pmatrix} A_{(1)}^{-1} = A \cdot A_{(1)}^{-1}$$

注意 $A_{(1)}^{-1}$ 是 $M_{\mathcal{B}}^{-1}$ 的前 m 行。设 $M_{\mathcal{B}}^{-1}$ 的前 m 行为矩阵 $C \in \mathbb{R}^{m \times n}$，则 C 满足：

$$CA = I_m, \quad CB = 0$$

这是因为 $M_{\mathcal{B}}^{-1} M_{\mathcal{B}} = I_n$ 的前 m 行给出：

$$C[A \mid B] = [CA \mid CB] = [I_m \mid 0]$$

因此，平行投影矩阵可写为：

$$P = AC$$

其中 $C \in \mathbb{R}^{m \times n}$ 满足 $CA = I_m$，$CB = 0$。

方法二：通过矩阵公式（代数方法）

设 $U, W \subseteq \mathbb{R}^n$，$\mathbb{R}^n = U \oplus W$，且：

- $A \in \mathbb{R}^{n \times m}$ 是 U 的基矩阵（$\mathrm{CS}(A) = U$，$\mathrm{rank}(A) = m$）
- $B \in \mathbb{R}^{m \times n}$ 是行满秩矩阵满足 $\mathrm{Null}(B) = W$（即 $B\mathbf{x} = \mathbf{0}$ 当且仅当 $\mathbf{x} \in W$）

则沿 W 方向到 U 的投影矩阵为：

$$P = A(BA)^{-1}B$$

详细推导：

1. **问题设定**：对于任意 $\mathbf{y} \in \mathbb{R}^n$，求其在直和分解 $\mathbb{R}^n = U \oplus W$ 下的 U 分量 $\mathbf{u} \in U$，使得 $\mathbf{y} = \mathbf{u} + \mathbf{w}$，其中 $\mathbf{w} \in W$。

2. **表示 U 分量**：由于 $\mathbf{u} \in U = \mathrm{CS}(A)$，存在系数向量 $\mathbf{x} \in \mathbb{R}^m$ 使得：

$$\mathbf{u} = A\mathbf{x}$$

3. **利用 W 分量性质**：由 $\mathbf{w} = \mathbf{y} - \mathbf{u} = \mathbf{y} - A\mathbf{x} \in W$，且 $W = \mathrm{Null}(B)$，得：

$$B(\mathbf{y} - A\mathbf{x}) = \mathbf{0}$$

即：

$$BA\mathbf{x} = B\mathbf{y}$$

4. **求解系数**：关键条件是 BA 可逆。这需要验证：

   - A 列满秩：$\mathrm{rank}(A) = m$
   - B 行满秩：$\mathrm{rank}(B) = m$
   - $\mathrm{Null}(B) = W$ 且 $\mathrm{CS}(A) = U$，由直和分解 $\mathbb{R}^n = U \oplus W$ 可保证 BA 是 $m \times m$ 可逆矩阵

因此：

$$\mathbf{x} = (BA)^{-1}B\mathbf{y}$$

5. **得到投影矩阵**：

$$\mathbf{u} = A\mathbf{x} = A(BA)^{-1}B\mathbf{y}$$

故投影矩阵为：

$$P = A(BA)^{-1}B$$

验证投影性质：

1. **幂等性**：

$$P^2 = A(BA)^{-1}B \cdot A(BA)^{-1}B = A(BA)^{-1}(BA)(BA)^{-1}B = A(BA)^{-1}B = P$$

2. **值域**：对于任意 $\mathbf{y} \in \mathbb{R}^n$，$P\mathbf{y} = A[(BA)^{-1}B\mathbf{y}] \in \mathrm{CS}(A) = U$。反之，对于任意 $\mathbf{u} \in U$，存在 $\mathbf{x}$ 使 $\mathbf{u} = A\mathbf{x}$，则：

$$P\mathbf{u} = A(BA)^{-1}B(A\mathbf{x}) = A(BA)^{-1}(BA)\mathbf{x} = A\mathbf{x} = \mathbf{u}$$

故 $\mathrm{Im}(P) = U$。

3. **核空间**：若 $\mathbf{y} \in \mathrm{Ker}(P)$，则 $A(BA)^{-1}B\mathbf{y} = \mathbf{0}$。由于 A 列满秩，有 $(BA)^{-1}B\mathbf{y} = \mathbf{0}$，进而 $B\mathbf{y} = \mathbf{0}$，故 $\mathbf{y} \in \mathrm{Null}(B) = W$。反之，若 $\mathbf{y} \in W$，则 $B\mathbf{y} = \mathbf{0}$，从而 $P\mathbf{y} = \mathbf{0}$。故 $\mathrm{Ker}(P) = W$。

例题 10.2 ($\mathbb{R}^3$ 到二维子空间 U 沿一维子空间 W 的平行投影)

已知 $U = Lin\left\{\begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}0\\1\\-1\end{pmatrix}\right\}$，$W = Lin\left\{\begin{pmatrix}2\\1\\0\end{pmatrix}\right\}$，求 $\mathbb{R}^3$ 到 U 沿 W 方向的投影矩阵。

方法一（基变换法）：

首先验证 $\mathbb{R}^3 = U \oplus W$：

- $\dim(U) = 2$，因为 $\begin{pmatrix}1\\0\\1\end{pmatrix}$ 与 $\begin{pmatrix}0\\1\\-1\end{pmatrix}$ 线性无关
- $\dim(W) = 1$
- 检查 $U \cap W = \{\mathbf{0}\}$：假设存在 α, β, γ 使得

$$\alpha\begin{pmatrix}1\\0\\1\end{pmatrix} + \beta\begin{pmatrix}0\\1\\-1\end{pmatrix} = \gamma\begin{pmatrix}2\\1\\0\end{pmatrix}$$

得方程组 $\alpha = 2\gamma, \beta = \gamma, \alpha - \beta = 0$，解得 $\alpha = \beta = \gamma = 0$

- $\dim(U) + \dim(W) = 2 + 1 = 3 = \dim(\mathbb{R}^3)$

取基 $\mathcal{B} = \left\{\begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}0\\1\\-1\end{pmatrix}, \begin{pmatrix}2\\1\\0\end{pmatrix}\right\}$，构造过渡矩阵：

$$M_{\mathcal{B}} = \begin{pmatrix}1 & 0 & 2\\0 & 1 & 1\\1 & -1 & 0\end{pmatrix}$$

在基 $\mathcal{B}$ 下，投影变换的矩阵为：

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

计算逆矩阵：

$$M_{\mathcal{B}}^{-1} = \begin{pmatrix} -1 & 2 & 2 \\ -1 & 2 & 1 \\ 1 & -1 & -1 \end{pmatrix}$$

标准基下的投影矩阵：

$$\begin{aligned} P &= M_{\mathcal{B}}[T]_{\mathcal{B}}^{\mathcal{B}} M_{\mathcal{B}}^{-1} \\ &= \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} -1 & 2 & 2 \\ -1 & 2 & 1 \\ 1 & -1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} -1 & 2 & 2 \\ -1 & 2 & 1 \\ 1 & -1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} -1 & 2 & 2 \\ -1 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

方法二（代数公式法）：

取 $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{pmatrix}$，满足 $CS(A) = U$。

构造 $B = \begin{pmatrix} 1 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$，满足 $N(B) = W$（验证：$B\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$）。

计算：

$$BA = \begin{pmatrix} 1 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix}$$

$$(BA)^{-1} = \frac{1}{-1+2} \begin{pmatrix} -1 & 2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 2 \\ -1 & 1 \end{pmatrix}$$

投影矩阵：

$$\begin{aligned} P &= A(BA)^{-1}B \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} -1 & 2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

$$
= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} -1 & 2 & 2 \\ -1 & 2 & 1 \end{pmatrix}
$$

$$
= \begin{pmatrix} -1 & 2 & 2 \\ -1 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix}
$$

例题 10.3 ($\mathbb{R}^3$ 到一维子空间 U 沿二维子空间 W 的平行投影) 已知 $U = Lin\left\{\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}\right\}$，$W = Lin\left\{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}\right\}$，求 $\mathbb{R}^3$ 到 U 沿 W 方向的投影矩阵。

方法一（基变换法）：

验证 $\mathbb{R}^3 = U \oplus W$：

- $\dim(U) = 1$
- $\dim(W) = 2$（$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ 与 $\begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}$ 线性无关）
- 检查 $U \cap W = \{\mathbf{0}\}$：假设存在 α, β, γ 使得

$$
\alpha \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \beta \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}
$$

得方程组 $\alpha = \beta, 0 = \beta + \gamma, \alpha = 3\gamma$，解得 $\alpha = \beta = \gamma = 0$

取基 $\mathcal{B} = \left\{\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}\right\}$，构造过渡矩阵：

$$
M_{\mathcal{B}} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 3 \end{pmatrix}
$$

在基 $\mathcal{B}$ 下，投影变换的矩阵为：

$$
[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}
$$

计算 $M_{\mathcal{B}}^{-1}$：

$$
\det(M_{\mathcal{B}}) = 1 \cdot (1 \cdot 3 - 1 \cdot 0) - 1 \cdot (0 \cdot 3 - 1 \cdot 1) = 3 - (-1) = 4
$$

$$M_{\mathcal{B}}^{-1} = \frac{1}{4}\begin{pmatrix} 3 & -3 & 1 \\ 1 & 3 & -1 \\ -1 & 1 & 1 \end{pmatrix}$$

标准基下的投影矩阵：

$$\begin{aligned} P &= M_{\mathcal{B}}[T]_{\mathcal{B}}^{\mathcal{B}} M_{\mathcal{B}}^{-1} \\ &= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \frac{1}{4}\begin{pmatrix} 3 & -3 & 1 \\ 1 & 3 & -1 \\ -1 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \cdot \frac{1}{4}\begin{pmatrix} 3 & -3 & 1 \\ 1 & 3 & -1 \\ -1 & 1 & 1 \end{pmatrix} \\ &= \frac{1}{4}\begin{pmatrix} 3 & -3 & 1 \\ 0 & 0 & 0 \\ 3 & -3 & 1 \end{pmatrix} \end{aligned}$$

方法二（代数公式法）：

取 $A = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$，满足 $CS(A) = U$。

构造 $B = (-3 \quad 3 \quad -1)$，满足 $N(B) = W$（验证：$B\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = 0$，$B\begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix} = 0$）。

计算：

$$BA = (-3, 3, -1)\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = -3 + 0 - 1 = -4$$

$$(BA)^{-1} = -\frac{1}{4}$$

投影矩阵：

$$\begin{aligned} P &= A(BA)^{-1}B \\ &= \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}\left(-\frac{1}{4}\right)(-3, 3, -1) \\ &= \frac{1}{4}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}(3, -3, 1) \\ &= \frac{1}{4}\begin{pmatrix} 3 & -3 & 1 \\ 0 & 0 & 0 \\ 3 & -3 & 1 \end{pmatrix} \end{aligned}$$

10.5 Orthogonal Projections

设 U 是实内积空间 V 的子空间，**正交投影**是 V 到 U 沿 $U^\perp$（U 的正交补）的投影。

定理 10.7 (正交投影的矩阵特征) *$n \times n$ 实矩阵 P 表示正交投影当且仅当 P 同时满足：*

1. *$P^2 = P$（幂等性）*
2. *$P^T = P$（对称性）*

对称性保证了投影方向与投影目标子空间正交，这正是“正交”投影的含义。

方法一：通过基变换

设 U 是 $\mathbb{R}^n$ 的 m 维子空间（$m < n$），$\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ 是 U 的一组标准正交基。设 $\mathcal{C} = \{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{v}_{m+1}, \ldots, \mathbf{v}_n\}$ 是 $\mathbb{R}^n$ 的一组标准正交基，其中 $\{\mathbf{v}_{m+1}, \ldots, \mathbf{v}_n\}$ 是 $U^\perp$ 的一组标准正交基。

在基 $\mathcal{C}$ 下，正交投影到 U 的变换矩阵非常简单：

$$[T]_{\mathcal{C}}^{\mathcal{C}} = \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix}_{n \times n}$$

其中 I_m 是 $m \times m$ 单位矩阵。

为得到标准基下的投影矩阵 P，需要进行基变换：

$$P = M_{\mathcal{C}} [T]_{\mathcal{C}}^{\mathcal{C}} M_{\mathcal{C}}^{-1}$$

其中 $M_{\mathcal{C}} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n]$ 是以 $\mathcal{C}$ 中向量为列构成的矩阵。

由于 $\mathcal{C}$ 是标准正交基，有 $M_{\mathcal{C}}^{-1} = M_{\mathcal{C}}^T$，因此：

$$P = M_{\mathcal{C}} \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} M_{\mathcal{C}}^T$$

将 $M_{\mathcal{C}}$ 分块为 $M_{\mathcal{C}} = [A \mid B]$,其中 $A = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_m] \in \mathbb{R}^{n \times m}$ 是 U 的基矩阵,$B = [\mathbf{v}_{m+1} \ \cdots \ \mathbf{v}_n] \in \mathbb{R}^{n \times (n-m)}$ 是 $U^\perp$ 的基矩阵，则：

$$P = [A \mid B] \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} A^T \\ B^T \end{pmatrix} = AA^T$$

这正是正交投影到 U 的标准公式。

方法二：通过矩阵公式

设 $U \subseteq \mathbb{R}^n$ 是 m 维子空间，$A \in \mathbb{R}^{n \times m}$ 是列满秩矩阵（$\operatorname{rank}(A) = m$），且 $\operatorname{CS}(A) = U$，即 A 的列张成 U。

正交投影到 W 的矩阵为：

$$P = A(A^T A)^{-1} A^T$$

详细推导：

设 $\mathbf{y} \in \mathbb{R}^n$，其在 U 上的正交投影 $\hat{\mathbf{y}}$ 是 U 中唯一满足 $\mathbf{y} - \hat{\mathbf{y}} \perp U$ 的向量。

由于 $\hat{\mathbf{y}} \in U = \operatorname{CS}(A)$，存在系数向量 $\mathbf{x} \in \mathbb{R}^m$ 使得 $\hat{\mathbf{y}} = A\mathbf{x}$。

正交条件 $\mathbf{y} - A\mathbf{x} \perp U$ 等价于 $\mathbf{y} - A\mathbf{x}$ 与 A 的所有列正交：

$$A^T(\mathbf{y} - A\mathbf{x}) = \mathbf{0}$$

即:

$$A^TA\mathbf{x} = A^T\mathbf{y}$$

由于 A 列满秩，A^TA 是 $m \times m$ 可逆矩阵，解得:

$$\mathbf{x} = (A^TA)^{-1}A^T\mathbf{y}$$

因此:

$$\hat{\mathbf{y}} = A\mathbf{x} = A(A^TA)^{-1}A^T\mathbf{y}$$

故投影矩阵为 $P = A(A^TA)^{-1}A^T$。

验证正交投影性质:

1. **幂等性**: $P^2 = A(A^TA)^{-1}A^T \cdot A(A^TA)^{-1}A^T = A(A^TA)^{-1}(A^TA)(A^TA)^{-1}A^T = A(A^TA)^{-1}A^T = P$

2. **对称性**: $P^T = [A(A^TA)^{-1}A^T]^T = A[(A^TA)^{-1}]^TA^T = A(A^TA)^{-1}A^T = P$

例题 10.4 ($\mathbb{R}^3$ 到一维子空间 U 的正交投影) 求 $\mathbb{R}^3$ 到子空间 $U = Lin\left\{\begin{pmatrix}2\\1\\1\end{pmatrix}\right\}$ 的正交投影矩阵。

方法一（基变换法）:

由于是正交投影，需构造 $\mathbb{R}^3$ 的一组标准正交基，其中包含 U 的标准正交基。首先，U 的基向量需要单位化:

$$\mathbf{u}_1 = \frac{1}{\left\|\begin{pmatrix}2\\1\\1\end{pmatrix}\right\|}\begin{pmatrix}2\\1\\1\end{pmatrix} = \frac{1}{\sqrt{6}}\begin{pmatrix}2\\1\\1\end{pmatrix}$$

需要找到 $U^\perp$ 的一组标准正交基。$U^\perp$ 由所有与 $\begin{pmatrix}2\\1\\1\end{pmatrix}$ 正交的向量组成:

$$U^\perp = \left\{\mathbf{x} \in \mathbb{R}^3 : 2x_1 + x_2 + x_3 = 0\right\}$$

选取 $U^\perp$ 的两个线性无关向量: $\begin{pmatrix}1\\-2\\0\end{pmatrix}$ 和 $\begin{pmatrix}0\\1\\-1\end{pmatrix}$。

这两个向量不正交，如果要构造标准正交基则需先用 *Gram-Schmidt* 正交化，好处是后面可以吃到正交矩阵的逆等于转置的便利。但也可以直接使用非正交基，成本就是需要老老实实求逆。我们这里直接求逆:

取基 $\mathcal{B} = \left\{\begin{pmatrix}2\\1\\1\end{pmatrix}, \begin{pmatrix}1\\-2\\0\end{pmatrix}, \begin{pmatrix}0\\1\\-1\end{pmatrix}\right\}$，构造过渡矩阵:

$$M_\mathcal{B} = \begin{pmatrix}2 & 1 & 0\\1 & -2 & 1\\1 & 0 & -1\end{pmatrix}$$

在基 $\mathcal{B}$ 下，正交投影到 U 的变换矩阵为：

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

计算 $M_{\mathcal{B}}^{-1}$：

$$M_{\mathcal{B}}^{-1} = \frac{1}{6}\begin{pmatrix} 2 & 1 & 1 \\ 2 & -2 & -2 \\ 2 & 1 & -5 \end{pmatrix}$$

标准基下的投影矩阵：

$$\begin{aligned} P &= M_{\mathcal{B}}[T]_{\mathcal{B}}^{\mathcal{B}}M_{\mathcal{B}}^{-1} \\ &= \begin{pmatrix} 2 & 1 & 0 \\ 1 & -2 & 1 \\ 1 & 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \frac{1}{6}\begin{pmatrix} 2 & 1 & 1 \\ 2 & -2 & -2 \\ 2 & 1 & -5 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \cdot \frac{1}{6}\begin{pmatrix} 2 & 1 & 1 \\ 2 & -2 & -2 \\ 2 & 1 & -5 \end{pmatrix} \\ &= \frac{1}{6}\begin{pmatrix} 4 & 2 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix} \end{aligned}$$

方法二（正交投影公式法）：

设 $A = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$，则 $CS(A) = U$。

计算：

$$A^TA = \begin{pmatrix} 2 & 1 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = 4 + 1 + 1 = 6$$

$$(A^TA)^{-1} = \frac{1}{6}$$

正交投影矩阵公式：

$$\begin{aligned} P &= A(A^TA)^{-1}A^T \\ &= \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \cdot \frac{1}{6} \cdot \begin{pmatrix} 2 & 1 & 1 \end{pmatrix} \\ &= \frac{1}{6}\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 2 & 1 & 1 \end{pmatrix} \end{aligned}$$

$$= \frac{1}{6}\begin{pmatrix} 4 & 2 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix}$$

10.6 Application of Orthogonal Projection: OLS

考虑一个线性变换把 $V = \mathbb{R}^n$ 正交投影到子空间 U 上（假设维度是 k），其正交投影矩阵是 P。请思考：对于任意一个 $\mathbf{v} \in V$，$P\mathbf{v} \in U$ 有什么特别的地方？

显然，$P\mathbf{v}$ 是 U 上离 $\mathbf{v}$ 最近的一个向量：

$$\forall \mathbf{u} \in U, \quad \|\mathbf{v} - P\mathbf{v}\| \leq \|\mathbf{v} - \mathbf{u}\|$$

这个性质非常宝贵，当我们希望最小化误差时，本质就是正交投影，从而获得最小的距离。下面我们来学习计量经济学的第一课——多元线性回归模型的 OLS 估计，来体会正交投影发挥的作用。

多元线性回归模型如下（标量形式）：

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, ..., n$$

我们可以把它写成矩阵形式：

$$Y = X\beta + \varepsilon$$

其中 Y 和 ε 都是 $n \times 1$ 向量，X 是一个 $n \times (k+1)$ 矩阵。

$$\hat{\varepsilon} = Y - X\hat{\beta}$$

(1). 代数方法：

OLS 希望最小化 $\hat{\varepsilon}'\hat{\varepsilon}$：

$$\hat{\varepsilon}'\hat{\varepsilon} = (Y - X\hat{\beta})'(Y - X\hat{\beta}) = (Y' - \hat{\beta}'X')(Y - X\hat{\beta}) = Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

$$FOC: \frac{\partial \hat{\varepsilon}'\hat{\varepsilon}}{\partial \hat{\beta}} = -X'Y - X'Y + 2X'X\hat{\beta} = 0$$

$$X'X\hat{\beta} = X'Y$$

如果 $X'X$ 非奇异（这等价于 X 必须是列满秩：$\text{rank}(X) = k+1$），则 OLS 存在解析解：

$$\boxed{\hat{\beta} = (X'X)^{-1}X'Y}$$

(2). 几何方法（正交投影）：

$$\hat{Y} = PY = X\hat{\beta}$$

P 是正交投影矩阵，把 Y 从 $V = \mathbb{R}^n$ 正交投影到子空间 $U = CS(X)$（维度是 $k+1 < n$）：

$$P = X(X'X)^{-1}X'$$

得到相同的答案：

$$\boxed{\hat{\beta} = (X'X)^{-1}X'Y}$$

注意 Y 的观测值可以被分成两部分：

$$Y = \hat{Y} + \hat{\varepsilon} = P_X Y + M_X Y$$

其中 P_X 是投影矩阵 (projection matrix)，又叫**帽子矩阵（hat matrix）**。M_X 是另一个投影矩阵，它把 Y 投影到 $U^\perp$，也就是 $CS(X)^\perp$，因此叫**湮灭矩阵（annihilator matrix）**或者叫化零矩阵。显然：

$$M_X = I_n - P_X$$

10.7 Orthogonal Projections in Abstract Vector Spaces

设 $(V, \langle\cdot,\cdot\rangle)$ 是一个内积空间，U 是 V 的有限维子空间。如果 $\{u_1, u_2, \ldots, u_n\}$ 是 U 的一组标准正交基，那么对于任意 $v \in V$，v 在 U 上的正交投影为：

$$P_U(v) = \sum_{i=1}^{n} \langle v, u_i \rangle u_i$$

考虑向量空间 $C[0,1]$（定义在区间 $[0,1]$ 上的连续函数），配备内积：

$$\langle f, g \rangle = \int_0^1 f(x)g(x)\,dx$$

设 $P_2[0,1] \subset C[0,1]$ 是所有次数不超过 2 的多项式构成的子空间。由以下三个函数构成 $P_2[0,1]$ 的一组标准正交基：

$$\begin{aligned}
h_1(x) &= 1 \\
h_2(x) &= \sqrt{3}(2x-1) \\
h_3(x) &= \sqrt{5}(6x^2-6x+1)
\end{aligned}$$

验证正交性

$$\begin{aligned}
\langle h_1, h_2 \rangle &= \int_0^1 1 \cdot \sqrt{3}(2x-1)\,dx = \sqrt{3}\left[x^2 - x\right]_0^1 = 0 \\
\langle h_1, h_3 \rangle &= \int_0^1 1 \cdot \sqrt{5}(6x^2-6x+1)\,dx = \sqrt{5}\left[2x^3 - 3x^2 + x\right]_0^1 = 0 \\
\langle h_2, h_3 \rangle &= \int_0^1 \sqrt{3}(2x-1) \cdot \sqrt{5}(6x^2-6x+1)\,dx \\
&= \sqrt{15}\int_0^1 (12x^3 - 18x^2 + 8x - 1)\,dx \\
&= \sqrt{15}\left[3x^4 - 6x^3 + 4x^2 - x\right]_0^1 = 0
\end{aligned}$$

验证单位长度

$$\begin{aligned}
\langle h_1, h_1 \rangle &= \int_0^1 1 \cdot 1\,dx = 1 \\
\langle h_2, h_2 \rangle &= \int_0^1 3(2x-1)^2\,dx = 3\int_0^1 (4x^2-4x+1)\,dx = 3\left[\frac{4}{3}x^3 - 2x^2 + x\right]_0^1 = 1 \\
\langle h_3, h_3 \rangle &= \int_0^1 5(6x^2-6x+1)^2\,dx = 5\int_0^1 (36x^4 - 72x^3 + 48x^2 - 12x + 1)\,dx \\
&= 5\left[\frac{36}{5}x^5 - 18x^4 + 16x^3 - 6x^2 + x\right]_0^1 = 1
\end{aligned}$$

求函数 $f(x) = \sqrt{x}$ 在 $P_2[0,1]$ 上的正交投影 $P(f)$。

根据正交投影公式：

$$P(f) = \langle f, h_1 \rangle h_1 + \langle f, h_2 \rangle h_2 + \langle f, h_3 \rangle h_3$$

计算内积：

$$
\begin{aligned}
\langle f, h_1 \rangle &= \int_0^1 \sqrt{x} \cdot 1 \, dx = \int_0^1 x^{1/2} \, dx = \left[\frac{2}{3} x^{3/2} \right]_0^1 = \frac{2}{3} \\
\langle f, h_2 \rangle &= \int_0^1 \sqrt{x} \cdot \sqrt{3}(2x-1) \, dx = \sqrt{3} \int_0^1 (2x^{3/2} - x^{1/2}) \, dx \\
&= \sqrt{3} \left[\frac{4}{5} x^{5/2} - \frac{2}{3} x^{3/2} \right]_0^1 = \sqrt{3} \left(\frac{4}{5} - \frac{2}{3} \right) = \sqrt{3} \cdot \frac{2}{15} = \frac{2\sqrt{3}}{15} \\
\langle f, h_3 \rangle &= \int_0^1 \sqrt{x} \cdot \sqrt{5}(6x^2 - 6x + 1) \, dx = \sqrt{5} \int_0^1 (6x^{5/2} - 6x^{3/2} + x^{1/2}) \, dx \\
&= \sqrt{5} \left[\frac{12}{7} x^{7/2} - \frac{12}{5} x^{5/2} + \frac{2}{3} x^{3/2} \right]_0^1 \\
&= \sqrt{5} \left(\frac{12}{7} - \frac{12}{5} + \frac{2}{3} \right) = \sqrt{5} \cdot \left(\frac{180 - 252 + 70}{105} \right) = \sqrt{5} \cdot \left(-\frac{2}{105} \right) = -\frac{2\sqrt{5}}{105}
\end{aligned}
$$

代入投影公式

$$
\begin{aligned}
P(f) &= \frac{2}{3} \cdot 1 + \frac{2\sqrt{3}}{15} \cdot \sqrt{3}(2x-1) - \frac{2\sqrt{5}}{105} \cdot \sqrt{5}(6x^2 - 6x + 1) \\
&= \frac{2}{3} + \frac{2}{15} \cdot 3(2x-1) - \frac{2}{105} \cdot 5(6x^2 - 6x + 1) \\
&= \frac{2}{3} + \frac{2}{5}(2x-1) - \frac{2}{21}(6x^2 - 6x + 1)
\end{aligned}
$$

化简结果

$$
\begin{aligned}
P(f) &= \frac{2}{3} + \frac{4}{5}x - \frac{2}{5} - \frac{12}{21}x^2 + \frac{12}{21}x - \frac{2}{21} \\
&= -\frac{12}{21}x^2 + \left(\frac{4}{5} + \frac{12}{21} \right) x + \left(\frac{2}{3} - \frac{2}{5} - \frac{2}{21} \right) \\
&= -\frac{4}{7}x^2 + \left(\frac{84}{105} + \frac{60}{105} \right) x + \left(\frac{70}{105} - \frac{42}{105} - \frac{10}{105} \right) \\
&= -\frac{4}{7}x^2 + \frac{144}{105}x + \frac{18}{105} \\
&= -\frac{4}{7}x^2 + \frac{48}{35}x + \frac{6}{35}
\end{aligned}
$$

因此，$\sqrt{x}$ 在 $P_2[0,1]$ 上的最佳二次多项式逼近为：

$$
P(f)(x) = -\frac{4}{7}x^2 + \frac{48}{35}x + \frac{6}{35}
$$

作为正交投影，$P(f)$ 是在 $P_2[0,1]$ 中最小化距离 $\|f - g\|$ 的函数，其中范数由内积诱导：

$$
\|f - g\|^2 = \langle f - g, f - g \rangle = \int_0^1 (f(x) - g(x))^2 \, dx
$$

因此，$P(f)$ 是 $\sqrt{x}$ 在最小二乘意义下的最佳二次多项式逼近，即它最小化：

$$
\int_0^1 \left(\sqrt{x} - (ax^2 + bx + c) \right)^2 \, dx
$$

对所有 $a, b, c \in \mathbb{R}$。

如果我们直接设 $g(x) = ax^2 + bx + c$，并最小化：

$$E(a,b,c) = \int_0^1 (\sqrt{x} - (ax^2 + bx + c))^2 \, dx$$

通过求偏导数 $\frac{\partial E}{\partial a} = \frac{\partial E}{\partial b} = \frac{\partial E}{\partial c} = 0$，可得正规方程：

$$\begin{aligned}
\int_0^1 x^4 \, dx \cdot a + \int_0^1 x^3 \, dx \cdot b + \int_0^1 x^2 \, dx \cdot c &= \int_0^1 x^{5/2} \, dx \\
\int_0^1 x^3 \, dx \cdot a + \int_0^1 x^2 \, dx \cdot b + \int_0^1 x \, dx \cdot c &= \int_0^1 x^{3/2} \, dx \\
\int_0^1 x^2 \, dx \cdot a + \int_0^1 x \, dx \cdot b + \int_0^1 1 \, dx \cdot c &= \int_0^1 x^{1/2} \, dx
\end{aligned}$$

计算积分：

$$\begin{pmatrix} \frac{1}{5} & \frac{1}{4} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \frac{2}{7} \\ \frac{2}{5} \\ \frac{2}{3} \end{pmatrix}$$

解此方程组将得到相同的 $a = -\frac{4}{7}$, $b = \frac{48}{35}$, $c = \frac{6}{35}$。

10.8 Exercises

11 Linear Transformations IV 线性变换 IV（从 $\mathbb{R}$ 推广至 $\mathbb{C}$）

11.1 Complex Vector Spaces

从本节开始，我们将把前面章节中所有基于实数域 $\mathbb{R}$ 的向量空间、矩阵、内积等概念推广到复数域 $\mathbb{C}$ 上。

复数集定义为：

$$\mathbb{C} = \{a + ib \mid a, b \in \mathbb{R}\}, \quad i = \sqrt{-1}.$$

设 $z = a + ib \in \mathbb{C}$：

- **复共轭**：$\bar{z} = a - ib$
- **模长**：$|z| = \sqrt{a^2 + b^2} = \sqrt{(a + ib)(a - ib)}$

共轭运算性质（$w, z \in \mathbb{C}$）：

$$\overline{w + z} = \bar{w} + \bar{z}, \quad \overline{wz} = \bar{w} \cdot \bar{z}, \quad \overline{\left(\frac{w}{z}\right)} = \frac{\bar{w}}{\bar{z}}.$$

模长运算性质：

$$|wz| = |w| \cdot |z|, \quad |w + z| \leq |w| + |z|.$$

设 $\mathbf{z} \in \mathbb{C}^n$：

$$\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}, \quad \bar{\mathbf{z}} = \begin{pmatrix} \bar{z}_1 \\ \bar{z}_2 \\ \vdots \\ \bar{z}_n \end{pmatrix}.$$

设 $A \in \mathbb{C}^{m \times n}$：

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \bar{A} = \begin{pmatrix} \bar{a}_{11} & \bar{a}_{12} & \cdots & \bar{a}_{1n} \\ \bar{a}_{21} & \bar{a}_{22} & \cdots & \bar{a}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{a}_{m1} & \bar{a}_{m2} & \cdots & \bar{a}_{mn} \end{pmatrix}.$$

若 A 为方阵，则：

$$|\bar{A}| = \overline{|A|}.$$

设 V 是复向量空间，V 上的一个**内积**是一个函数 $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$，满足：

1. $\langle \alpha u + \beta v, w \rangle = \alpha \langle u, w \rangle + \beta \langle v, w \rangle$（对第一个变量线性）
2. $\langle u, v \rangle = \overline{\langle v, u \rangle}$（共轭对称性）
3. $\langle u, u \rangle \geq 0$，且 $\langle u, u \rangle = 0 \iff u = 0$（正定性）

注意：复内积不是双线性的，而是对第一个变量线性、对第二个变量共轭线性。

在 $\mathbb{C}^n$ 中，标准内积定义为：

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \bar{\mathbf{v}} = u_1 \bar{v}_1 + u_2 \bar{v}_2 + \cdots + u_n \bar{v}_n.$$

例如在 $\mathbb{C}^2$ 中：

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = u_1 \overline{v_1} + u_2 \overline{v_2}.$$

11.2 Special Complex Matrices

设 $A \in \mathbb{C}^{m\times n}$，其**共轭转置 (Hermitian transpose)**（又称 Hermitian 转置）定义为：

$$A^* = \bar{A}^T = \overline{A^T}.$$

若 A 为实矩阵，则 $A^* = A^T$。

Hermitian/共轭转置是普通转置在复数域上的自然推广。

- $(A^*)^* = A$
- $(AB)^* = B^*A^*$
- $|A^*| = \overline{|A|}$
- 对 $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$，标准内积可写为：
$$\langle \mathbf{u}, \mathbf{v}\rangle = \mathbf{v}^*\mathbf{u}$$
- 伴随性质（对 $A \in \mathbb{C}^{n\times n}$）：
$$\langle A\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, A^*\mathbf{y}\rangle, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{C}^n.$$

若 $A \in \mathbb{C}^{n\times n}$ 满足 $A^* = A$，则称 A 为 **Hermitian 矩阵**。

Hermitian 矩阵是实对称矩阵在复域上的推广：

- 主对角线元素必为实数
- 非对角线元素满足 $a_{ij} = \overline{a_{ji}}$

例如：

$$\begin{pmatrix} 1 & i \\ -i & 2 \end{pmatrix}, \quad \begin{pmatrix} 3 & 4 & 7 \\ 4 & 2 & 5 \\ 7 & 5 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1+2i & 7 \\ 1-2i & 2 & -i \\ 7 & i & 3 \end{pmatrix}.$$

1. 所有特征值均为实数
2. 对应不同特征值的特征向量相互正交（在标准内积下）

若 $A \in \mathbb{C}^{n\times n}$ 满足 $A^*A = I_n$，则称 A 为 **酉矩阵 (unitary matrix)**。

酉矩阵是实正交矩阵在复域上的推广。下列命题等价：

1. A 是酉矩阵
2. $A^* = A^{-1}$
3. $AA^* = I_n$
4. A 的列向量构成 $\mathbb{C}^n$ 的一组标准正交基（在复内积意义下）

酉矩阵的性质：

- $|\det A| = 1$，故 A 可逆
- 所有特征值的模均为 1：$|\lambda| = 1$

若 $A\in\mathbb{C}^{n\times n}$ 满足 $A^*A=AA^*$，则称 A 为 **正规矩阵 (normal matrix)**。

- 所有 Hermitian 矩阵都是正规矩阵（因为 $A^*=A$ 时 $A^*A=AA^*=A^2$）
- 所有酉矩阵也都是正规矩阵（因为 $A^*A=AA^*=I_n$）
- 存在既不是 Hermitian 也不是酉矩阵的正规矩阵

若矩阵 A 同时满足 $A^*=A$ 且 $A^*A=I$，则它既是 Hermitian 又是酉矩阵。例如：

$$\begin{pmatrix}\cos\theta & \sin\theta\\ \sin\theta & -\cos\theta\end{pmatrix},\quad \begin{pmatrix}\cos\theta & i\sin\theta\\ -i\sin\theta & -\cos\theta\end{pmatrix},\quad I_n,\quad -I_n.$$

定理 11.1 (复正规矩阵的酉对角化) *设 $A\in\mathbb{C}^{n\times n}$，则下列命题等价：*

1. *A 是正规矩阵：$A^*A=AA^*$*
2. *A 可被酉矩阵对角化：存在酉矩阵 U 与对角矩阵 D 使得*

$$A=UDU^*,$$

其中 D 的对角元素为 A 的特征值。

3. *$\mathbb{C}^n$ 有一组由 A 的特征向量构成的标准正交基。*

特别地：

- Hermitian 矩阵可被酉对角化为实对角矩阵（特征值全为实数）
- 酉矩阵可被酉对角化为对角矩阵，其对角元模长为 1

表 3: 实数域与复数域上转置、内积与矩阵类型的对应关系

实数域 $\mathbb{R}$	**复数域** $\mathbb{C}$
转置 A^T	**共轭转置** A^*
对称矩阵 $A^T=A$	**Hermitian 矩阵** $A^*=A$
正交矩阵 $A^TA=I$	**酉矩阵** $A^*A=I$
标准内积 $\mathbf{u}^T\mathbf{v}$	标准内积 $\mathbf{u}^*\mathbf{v}$
谱定理：实对称矩阵可正交对角化	谱定理：Hermitian 矩阵可酉对角化

11.3 Unitary Diagonalisation

我们先把前面学的实对角化推广到复数域，使上一章的内容成为这一节的特例。
设 $A\in\mathbb{C}^{n\times n}$，若存在 $\lambda\in\mathbb{C}$ 和非零向量 $\mathbf{x}\in\mathbb{C}^n$ 使得

$$A\mathbf{x}=\lambda\mathbf{x},$$

则称 λ 为 A 的**特征值**，$\mathbf{x}$ 为对应的**特征向量**。

特征值可通过求解特征方程得到：

$$\det(A-\lambda I_n)=0.$$

若特征多项式可分解为：

$$\det(A-\lambda I_n)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_k)^{m_k},$$

其中 $m_1+m_2+\cdots+m_k=n$，则：

- λ_i 为 A 的特征值
- $a_{\lambda_i}=m_i$ 称为 λ_i 的**代数重数**

对特征值 λ，其**特征空间**定义为：

$$E(A,\lambda)=\text{null}(A-\lambda I_n).$$

该空间的维数称为 λ 的**几何重数**：

$$g_\lambda=\dim E(A,\lambda).$$

几何重数表示对应 λ 的线性无关特征向量的最大个数。总满足：

$$1\leq g_\lambda\leq a_\lambda.$$

称 A 可对角化，若存在可逆矩阵 P 使得：

$$P^{-1}AP=D,$$

其中 D 为对角矩阵。

定理 11.2 (可对角化等价条件) 对 $A\in\mathbb{C}^{n\times n}$，以下命题等价：

1. A 可对角化
2. A 有 n 个线性无关的特征向量（构成 $\mathbb{C}^n$ 的一组基）
3. 对每个特征值 λ，其几何重数等于代数重数：$g_\lambda=a_\lambda$

命题 11.1 (实矩阵的共轭特征对) 若 $A\in\mathbb{R}^{n\times n}$ 且 λ 是 A 的复特征值，对应特征向量为 $\mathbf{v}$，则：

- $\bar{\lambda}$ 也是 A 的特征值
- $\bar{\mathbf{v}}$ 是对应 $\bar{\lambda}$ 的特征向量

称 $A\in\mathbb{C}^{n\times n}$ 可**酉对角化**，若存在酉矩阵 U 使得：

$$U^*AU=D\quad(\text{或等价地}A=UDU^*),$$

其中 D 为对角矩阵。酉对角化是实数域中正交对角化的复推广。

定理 11.3 (酉对角化等价条件) 对 $A\in\mathbb{C}^{n\times n}$，以下命题等价：

1. A 可酉对角化
2. A 有 n 个标准正交的特征向量（构成 $\mathbb{C}^n$ 的一组标准正交基）
3. A 是正规矩阵：$A^*A=AA^*$

推论 11.1 (实对称矩阵的正交对角化) 设 $A\in\mathbb{R}^{n\times n}$ 为实对称矩阵，则：

- A 的所有特征值均为实数
- A 可正交对角化：存在正交矩阵 Q 使得 $Q^TAQ=D$

此为正交代数基本定理在实数域的特例。

11.4 Spectral Decompositions

设 $\mathbf{v} \in \mathbb{C}^n$ 是单位向量（$\|\mathbf{v}\| = 1$），则矩阵 $P = \mathbf{v}\mathbf{v}^*$ 是到 $\mathrm{span}\{\mathbf{v}\}$ 的**正交投影算子**，满足：

- $P^2 = P$（幂等性）
- $P^* = P$（自伴性，Hermitian）
- 对任意 $\mathbf{x} \in \mathbb{C}^n$，$P\mathbf{x} = \langle \mathbf{x}, \mathbf{v} \rangle \mathbf{v}$

设 $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ 是 $\mathbb{C}^n$ 的一组标准正交基，则单位矩阵可分解为：

$$I_n = \mathbf{v}_1\mathbf{v}_1^* + \mathbf{v}_2\mathbf{v}_2^* + \cdots + \mathbf{v}_n\mathbf{v}_n^*,$$

其中 $\mathbf{v}_i\mathbf{v}_i^*$ 是秩为 1 的正交投影矩阵。

证明： 对任意 $\mathbf{x} \in \mathbb{C}^n$，由于 $\{\mathbf{v}_i\}$ 是基，有

$$\mathbf{x} = \sum_{i=1}^n \langle \mathbf{x}, \mathbf{v}_i \rangle \mathbf{v}_i = \sum_{i=1}^n (\mathbf{v}_i^*\mathbf{x})\mathbf{v}_i = \sum_{i=1}^n (\mathbf{v}_i\mathbf{v}_i^*)\mathbf{x}.$$

因此

$$I_n\mathbf{x} = \left(\sum_{i=1}^n \mathbf{v}_i\mathbf{v}_i^* \right) \mathbf{x},$$

对所有 $\mathbf{x}$ 成立，故

$$I_n = \sum_{i=1}^n \mathbf{v}_i\mathbf{v}_i^* = \sum_{i=1}^n P_i.$$

设 A 是 $n \times n$ 正规矩阵。由复谱定理，存在酉矩阵 U 使得

$$A = UDU^*,$$

其中 $D = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$，$\lambda_i$ 是 A 的特征值，U 的列向量 $\mathbf{u}_i$ 是对应的标准正交特征向量。

推导：

$$\begin{aligned}
A = UDU^* &= U \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} U^* \\
&= U \left(\sum_{i=1}^n \lambda_i \mathbf{e}_i \mathbf{e}_i^T \right) U^* \quad (\text{将 } D \text{ 写为投影的和}) \\
&= \sum_{i=1}^n \lambda_i (U\mathbf{e}_i)(U\mathbf{e}_i)^* \quad (\text{因 } (AB)^* = B^*A^*) \\
&= \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^*.
\end{aligned}$$

其中 $\mathbf{e}_i$ 是第 i 个标准基向量，且 $\mathbf{u}_i = U\mathbf{e}_i$ 是 U 的第 i 列。

定理 11.4 (谱分解) *若 A 是正规矩阵，且 $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ 是 A 的一组标准正交特征向量，对应特征值为 $\lambda_1, \lambda_2, \ldots, \lambda_n$，则 A 的***谱分解***为：*

$$A = \lambda_1\mathbf{v}_1\mathbf{v}_1^* + \lambda_2\mathbf{v}_2\mathbf{v}_2^* + \cdots + \lambda_n\mathbf{v}_n\mathbf{v}_n^*.$$

谱分解 $A = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^*$ 表明：

- 在由特征向量 $\mathbf{u}_i$ 张成的方向上，A 的作用相当于数乘 λ_i
- 任意向量 $\mathbf{x}$ 可分解为 $\mathbf{x} = \sum \langle \mathbf{x}, \mathbf{u}_i \rangle \mathbf{u}_i$
- $A\mathbf{x} = \sum \lambda_i \langle \mathbf{x}, \mathbf{u}_i \rangle \mathbf{u}_i$

若 A 有**互不相同的特征值** $\lambda_1, \ldots, \lambda_k$，对应特征空间 $E_1, \ldots, E_k$，令 P_i 为到 E_i 的正交投影，则谱分解可写为：

$$A = \sum_{i=1}^{k} \lambda_i P_i,$$

其中：

- $P_i P_j = 0$ 当 $i \neq j$（正交性）
- $P_i^2 = P_i$（幂等性）
- $\sum_{i=1}^{k} P_i = I_n$（单位分解）

这种分解是唯一的，称为 A 的**谱定理**。

例题 11.1 设 $A = \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}$（*Hermitian* 矩阵）：

1. *特征值：$\lambda_1 = 0, \lambda_2 = 2$*
2. *对应归一化特征向量：*

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

3. *谱分解：*

$$\begin{aligned} A &= 0 \cdot \mathbf{u}_1 \mathbf{u}_1^* + 2 \cdot \mathbf{u}_2 \mathbf{u}_2^* = 2 \cdot \frac{1}{2} \begin{pmatrix} 1 \\ -i \end{pmatrix} \begin{pmatrix} 1 & i \end{pmatrix} \\ &= \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} \end{aligned}$$

例题 11.2 *考虑实对称矩阵*

$$A = \begin{pmatrix} 7 & 0 & 9 \\ 0 & 2 & 0 \\ 9 & 0 & 7 \end{pmatrix}.$$

1. *求正交矩阵 P 使得 $P^T AP$ 为对角矩阵；*
2. *将 A 写成谱分解形式 $A = \lambda_1 E_1 + \lambda_2 E_2 + \lambda_3 E_3$，其中 λ_i 是 A 的特征值，$E_i = \mathbf{v}_i \mathbf{v}_i^T$（$\mathbf{v}_i$ 是对应的单位特征向量）；*
3. *验证：当 $i \neq j$ 时 $E_i E_j = O$，且 $E_i^2 = E_i$。*

第一步：求特征值与特征向量通过观察或计算，可发现 A 的特征值与对应的特征向量为：

- 特征值 $\lambda_1 = 2$，特征向量 $\mathbf{v}_1 = (0, 1, 0)^T$
- 特征值 $\lambda_2 = 16$，特征向量 $\mathbf{v}_2 = (1, 0, 1)^T$

- 特征值 $\lambda_3 = -2$，特征向量 $\mathbf{v}_3 = (1, 0, -1)^T$

容易验证这些特征向量两两正交。

第二步：单位化并构造正交矩阵将特征向量单位化：

$$\mathbf{u}_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{u}_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

构造正交矩阵 P 以这些单位特征向量为列：

$$P = \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

则有

$$P^T A P = D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

第三步：谱分解定义投影矩阵：

$$E_1 = \mathbf{u}_1 \mathbf{u}_1^T = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} (0, 1, 0) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$E_2 = \mathbf{u}_2 \mathbf{u}_2^T = \frac{1}{2} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} (1, 0, 1) = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix},$$

$$E_3 = \mathbf{u}_3 \mathbf{u}_3^T = \frac{1}{2} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} (1, 0, -1) = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

则 A 的谱分解为：

$$A = \lambda_1 E_1 + \lambda_2 E_2 + \lambda_3 E_3 = 2E_1 + 16E_2 - 2E_3.$$

验证：

$$\begin{aligned} 2E_1 + 16E_2 - 2E_3 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 8 & 0 & 8 \\ 0 & 0 & 0 \\ 8 & 0 & 8 \end{pmatrix} - \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 7 & 0 & 9 \\ 0 & 2 & 0 \\ 9 & 0 & 7 \end{pmatrix} = A. \end{aligned}$$

第四步：验证投影矩阵的性质

- 幂等性（$E_i^2 = E_i$）：

$$E_1^2 = E_1, \quad E_2^2 = E_2, \quad E_3^2 = E_3.$$

这是因为对任意单位向量 $\mathbf{u}$，有

$$(\mathbf{u}\mathbf{u}^T)^2 = \mathbf{u}(\mathbf{u}^T\mathbf{u})\mathbf{u}^T = \mathbf{u}\mathbf{u}^T.$$

- 正交性（$i \neq j$ 时 $E_iE_j = O$）：

$$E_1E_2 = O, \quad E_1E_3 = O, \quad E_2E_3 = O.$$

这是因为当 $\mathbf{u}_i \perp \mathbf{u}_j$ 时，$\mathbf{u}_i^T\mathbf{u}_j = 0$，从而

$$(\mathbf{u}_i\mathbf{u}_i^T)(\mathbf{u}_j\mathbf{u}_j^T) = \mathbf{u}_i(\mathbf{u}_i^T\mathbf{u}_j)\mathbf{u}_j^T = O.$$

11.5 Exercises

12 Linear Transformations V 线性变换 V（从方阵推广至任意矩阵）

12.1 Singular Values

前文讨论的特征值分解（对角化、谱分解）要求矩阵是**方阵**且满足一定条件（如正规、可对角化）。然而在实际应用中，我们常遇到非方阵（如数据矩阵、线性变换在不同维空间之间）。为了将“特征值-特征向量”的思想推广到任意 $m \times n$ 复矩阵，我们引入**奇异值分解 (SVD)**。

SVD 的核心思想是：

- 对任意矩阵 $A \in \mathbb{C}^{m\times n}$，考虑两个 Hermitian 半正定矩阵：$A^*A \in \mathbb{C}^{n\times n}$ 和 $AA^* \in \mathbb{C}^{m\times m}$
- 这两个矩阵有相同的非零特征值，其平方根称为 A 的**奇异值**
- 将 A 分解为 $A = U\Sigma V^*$，其中 U, V 为酉矩阵，Σ 为“对角”矩阵（奇异值矩阵）

命题 12.1 (A^*A 与 AA^* 的性质) 设 $A \in \mathbb{C}^{m\times n}$，则：

1. A^*A 与 AA^* 都是 *Hermitian* 矩阵（$(A^*A)^* = A^*A$）

2. 两者特征值均为实数

3. 两者特征值均非负（半正定性）

4. A^*A 与 AA^* 有**相同的非零特征值**

证明（非负性）：设 λ 是 A^*A 的特征值，对应特征向量 $\mathbf{u}$（$\|\mathbf{u}\| = 1$），则

$$\lambda = \lambda\mathbf{u}^*\mathbf{u} = \mathbf{u}^*(\lambda\mathbf{u}) = \mathbf{u}^*(A^*A\mathbf{u}) = (A\mathbf{u})^*(A\mathbf{u}) = \|A\mathbf{u}\|^2 \geq 0.$$

证明（相同非零特征值）：若 $A^*A\mathbf{u} = \lambda\mathbf{u}$（$\lambda > 0$），则

$$AA^*(A\mathbf{u}) = A(A^*A\mathbf{u}) = \lambda(A\mathbf{u}),$$

且 $A\mathbf{u} \neq 0$（否则 $\lambda = 0$），故 λ 也是 AA^* 的特征值。

定义 12.1 (奇异值) 设 $A \in \mathbb{C}^{m\times n}$，$A^*A$ 的非零特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$，则 A 的**奇异值**为

$$\sigma_i = \sqrt{\lambda_i}, \quad i = 1, \ldots, k.$$

通常按降序排列：$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$。

例题 12.1 求矩阵 $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$ 的奇异值。

解：计算 AA^*：

$$AA^* = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

特征方程：

$$\det(AA^* - \lambda I) = \begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = (\lambda - 3)(\lambda - 1) = 0.$$

特征值 $\lambda_1 = 3, \lambda_2 = 1$，故奇异值为 $\sigma_1 = \sqrt{3}, \sigma_2 = 1$。

12.2 Singular Value Decomposition (SVD)

设 A^*A 的非零特征值为 $\lambda_1,\ldots,\lambda_k$，对应标准正交特征向量 $\mathbf{u}_1,\ldots,\mathbf{u}_k\in\mathbb{C}^n$。定义

$$\mathbf{v}_i=\frac{1}{\sqrt{\lambda_i}}A\mathbf{u}_i=\frac{1}{\sigma_i}A\mathbf{u}_i\in\mathbb{C}^m,\quad i=1,\ldots,k.$$

则 $\{\mathbf{v}_1,\ldots,\mathbf{v}_k\}$ 是 AA^* 对应特征值 λ_i 的标准正交特征向量集。

验证正交性：

$$\langle\mathbf{v}_i,\mathbf{v}_j\rangle=\frac{1}{\sqrt{\lambda_i\lambda_j}}(A\mathbf{u}_i)^*(A\mathbf{u}_j)=\frac{1}{\sqrt{\lambda_i\lambda_j}}\mathbf{u}_i^*(A^*A\mathbf{u}_j)=\frac{\lambda_j}{\sqrt{\lambda_i\lambda_j}}\mathbf{u}_i^*\mathbf{u}_j=\delta_{ij}.$$

定理 12.1 (奇异值分解 (SVD)) *设 $A\in\mathbb{C}^{m\times n}$，秩为 r（即非零奇异值个数），则存在酉矩阵 $U\in\mathbb{C}^{m\times m}$ 和 $V\in\mathbb{C}^{n\times n}$，以及对角矩阵 $\Sigma\in\mathbb{R}^{m\times n}$（$m\times n$ 的“对角”矩阵），使得*

$$A=U\Sigma V^*,$$

其中

$$\Sigma=\begin{pmatrix}\Sigma_r & 0\\ 0 & 0\end{pmatrix},\quad \Sigma_r=\mathrm{diag}(\sigma_1,\sigma_2,\ldots,\sigma_r),\quad \sigma_1\ge\sigma_2\ge\cdots\ge\sigma_r>0.$$

将 U,V 按列分块：

- $V=[\mathbf{v}_1\cdots\mathbf{v}_r\mid\mathbf{v}_{r+1}\cdots\mathbf{v}_n]$，前 r 列为 A^*A 的非零特征值对应特征向量（右奇异向量）
- $U=[\mathbf{u}_1\cdots\mathbf{u}_r\mid\mathbf{u}_{r+1}\cdots\mathbf{u}_m]$，前 r 列为 AA^* 的非零特征值对应特征向量（左奇异向量），由 $\mathbf{u}_i=\frac{1}{\sigma_i}A\mathbf{v}_i$ 得到
- 则有 $A\mathbf{v}_i=\sigma_i\mathbf{u}_i$（$i=1,\ldots,r$）

SVD 可写为秩-1 矩阵的和：

$$A=\sum_{i=1}^{r}\sigma_i\mathbf{u}_i\mathbf{v}_i^*.$$

这就是**奇异值分解的外积形式**。

例题 12.2 *继续前面的例子，求 $A=\begin{pmatrix}1&1&0\\1&0&1\end{pmatrix}$ 的 SVD。*

已得：奇异值 $\sigma_1=\sqrt{3},\sigma_2=1$。

步骤 *1*：求 A^*A 的特征向量

$$A^*A=\begin{pmatrix}1&1\\1&0\\0&1\end{pmatrix}\begin{pmatrix}1&1&0\\1&0&1\end{pmatrix}=\begin{pmatrix}2&1&1\\1&1&0\\1&0&1\end{pmatrix}.$$

特征值：由 AA^* 知 A^*A 有特征值 3, 1, 0。

- *对 $\lambda=3$：解 $(A^*A-3I)\mathbf{v}=0$，得 $\mathbf{v}_1=\frac{1}{\sqrt{3}}(1,1,1)^T$*
- *对 $\lambda=1$：解 $(A^*A-I)\mathbf{v}=0$，得 $\mathbf{v}_2=\frac{1}{\sqrt{2}}(1,0,-1)^T$*
- *对 $\lambda=0$：解 $A^*A\mathbf{v}=0$，得 $\mathbf{v}_3=\frac{1}{\sqrt{6}}(1,-2,1)^T$*

步骤 *2*：求左奇异向量

$$\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2} A\mathbf{v}_2 = \frac{1}{1} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

标准化：$\mathbf{u}_1 = \frac{1}{\sqrt{2}}(1,1)^T, \mathbf{u}_2 = (1,0)^T$（已标准）。

步骤 *3*：构造矩阵

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & 1 \\ \frac{1}{\sqrt{2}} & 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix}.$$

验证：$A = U\Sigma V^*$。

外积形式：

$$A = \sqrt{3} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} + 1 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & -1 \end{pmatrix}.$$

下面我们来讨论一下奇异值分解的几何意义。设 $A \in \mathbb{R}^{m\times n}$，SVD $A = U\Sigma V^T$ 表示：

1. V^T：在 $\mathbb{R}^n$ 中的旋转/反射
2. Σ：沿坐标轴的伸缩（奇异值为伸缩因子）
3. U：在 $\mathbb{R}^m$ 中的旋转/反射

因此，A 将单位球面映射为椭球面，其半轴长等于奇异值。

12.3 Eckart–Young Theorem and Low-rank Approximation

SVD 的外积形式：

$$A = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^*.$$

可以给到我们一个直接的启示：在秩不超过 k 的所有矩阵中，最接近 A（在 Frobenius 范数下）的是：

$$A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i'$$

近似误差：

$$\|A - A_k\|_F^2 = \sum_{i=k+1}^{r} \sigma_i^2$$

如果我们只保留前 k 个奇异值，舍弃的平方和就是后面奇异值的平方和。为了最小化误差，自然要保留最大的那些奇异值，因为它们携带最多的平方“能量”。

事实上，我们有如下定理：

定理 12.2 (Eckart–Young 定理) 设 $A \in \mathbb{R}^{m\times n}$（或 $\mathbb{C}^{m\times n}$）是一个矩阵，其奇异值分解为

$$A = U\Sigma V^*, \quad \Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p), \quad p = \min(m, n),$$

其中 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$，$U$ 和 V 为酉矩阵（正交矩阵），V^* 表示 V 的共轭转置。

对于任意整数 k 满足 $0 \leq k \leq p$，定义秩不超过 k 的矩阵集合

$$\mathcal{M}_k = \{B \in \mathbb{R}^{m\times n} : \mathrm{rank}(B) \leq k\}.$$

记

$$A_k = U\Sigma_k V^*, \quad \Sigma_k = \mathrm{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0).$$

则以下结论成立：

1. ***Frobenius* 范数最优性**：

$$\min_{B\in\mathcal{M}_k} \|A - B\|_F = \|A - A_k\|_F = \sqrt{\sum_{i=k+1}^{p} \sigma_i^2}.$$

且 A_k 是达到该最小值的一个矩阵（若 $\sigma_k > \sigma_{k+1}$，则唯一）。

2. **谱范数（*2*-范数）最优性**：

$$\min_{B\in\mathcal{M}_k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1},$$

其中约定 $\sigma_{p+1} = 0$。

更一般地，对于任意酉不变范数 $\|\cdot\|$，A_k 是 $\min_{\mathrm{rank}(B)\leq k} \|A - B\|$ 的一个解。

注 12.1 对于 *Frobenius* 范数和谱范数，该定理最早由 *Eckart* 与 *Young (1936)* 证明。后来 *Mirsky (1960)* 将其推广到所有酉不变范数，因此有时也称为 *Eckart–Young–Mirsky* 定理。

保留奇异值最大的 k 个谱来使该矩阵为最接近原矩阵的秩-k 矩阵的想法是自然的，毕竟奇异值越大，变换在该方向上“放大”输入的能力越强，该方向对输出的贡献越大，代表越主要的成分。但范数是可以选择的，于是问题紧而就变成了：哪些范数不会破坏这个结果。Eckart–Young 定理的主要贡献就在于给出了这个问题的答案。

下面我们就可以回到之前留下的问题：低秩近似，来给出解答。

例题 12.3 (低秩近似的解析解) 给定一个大矩阵 A（例如 $m \times n, m, n$ 很大），想找一个秩 r 较小的矩阵 B 来近似 A，使得误差 $\|A - B\|$ 尽可能小。

方法：使用奇异值分解（*SVD*），取前 r 个最大的奇异值及其对应的奇异向量：

$$A \approx U_r \Sigma_r V_r'$$

这是 *Eckart-Young* 定理保证的在 *Frobenius* 范数或谱范数下的最佳秩 r 近似。

低秩近似有许多重要应用，例如图像压缩、推荐系统、去噪。图像矩阵的无损保存需要 mn 个数，但如果用低秩近似表示，只需要存储 $r(m + n + 1)$ 个数。

12.4 Left and Right Inverses

对于方阵 $A \in \mathbb{C}^{n\times n}$，若 $\det A \neq 0$，则存在唯一的**逆矩阵** A^{-1} 满足：

$$AA^{-1} = A^{-1}A = I_n.$$

然而当 A 为一般 $m \times n$ 矩阵时，上述定义失效，因为：

- A 可能不是方阵（$m \neq n$）
- 即使 A 是方阵，也可能不可逆（奇异）

我们需要推广"逆"的概念，使之适用于任意矩阵。

定义 12.2 (左逆与右逆) *设 $A \in \mathbb{C}^{m\times n}$。*

- *若存在 $L \in \mathbb{C}^{n\times m}$ 使得 $LA = I_n$，则称 L 为 A 的***左逆***（left inverse）*
- *若存在 $R \in \mathbb{C}^{n\times m}$ 使得 $AR = I_m$，则称 R 为 A 的***右逆***（right inverse）*

定理 12.3 (左逆存在条件) *A 有左逆当且仅当 A 是***列满秩***（full column rank）的，即 $\mathrm{rank}(A) = n \leq m$。此时左逆为*

$$L = (A^*A)^{-1}A^*.$$

定理 12.4 (右逆存在条件) *A 有右逆当且仅当 A 是***行满秩***(full row rank)的，即 $\mathrm{rank}(A) = m \leq n$。此时右逆为*

$$R = A^*(AA^*)^{-1}.$$

注：

- 左逆不唯一（除非 $m = n$）；右逆也不唯一（除非 $m = n$）
- 当 $m = n$ 且 A 可逆时，左逆 = 右逆 = A^{-1}

12.5 Generalised Inverses

在线性代数中，广义逆可根据满足条件的不同分为多个等级，其中最重要的两个类别是**弱广义逆**（只满足部分条件）和**强广义逆**（满足全部四个条件）。

定义 12.3 (弱广义逆（或 {1}-逆）) *设 $A \in \mathbb{C}^{m\times n}$，若矩阵 $A^- \in \mathbb{C}^{n\times m}$ 满足*

$$AA^-A = A,$$

*则称 A^- 为 A 的***一个弱广义逆***，记作 $A\{1\}$。*

- 对任意矩阵 A，弱广义逆总是存在的（不唯一）
- 弱广义逆的集合记作 $A\{1\} = \{X \mid AXA = A\}$
- 若 A 可逆，则 $A^{-1} \in A\{1\}$ 且是唯一的
- 弱广义逆可用于求解**相容线性方程组**

定理 12.5 (相容方程组的解) 设 $A\mathbf{x}=\mathbf{b}$ 是相容方程组（即有解），则对任意弱广义逆 A^-，

$$\mathbf{x}=A^-\mathbf{b}+(I-A^-A)\mathbf{z},\quad \forall\mathbf{z}\in\mathbb{C}^n$$

给出方程的所有解。

定义 12.4 ($\{1,2\}$-逆) 若 X 同时满足：

1. $AXA=A$ （$\{1\}$-条件）
2. $XAX=X$ （$\{2\}$-条件：幂等性）

则称 X 为 A 的一个 $\{1,2\}$-逆。

定义 12.5 ($\{1,3\}$-逆) 若 X 同时满足：

1. $AXA=A$ （$\{1\}$-条件）
2. $(AX)^*=AX$ （$\{3\}$-条件：AX 为 *Hermitian*）

则称 X 为 A 的一个 $\{1,3\}$-逆。

定义 12.6 ($\{1,4\}$-逆) 若 X 同时满足：

1. $AXA=A$ （$\{1\}$-条件）
2. $(XA)^*=XA$ （$\{4\}$-条件：XA 为 *Hermitian*）

则称 X 为 A 的一个 $\{1,4\}$-逆。

定义 12.7 (M-P 广义逆 = $\{1,2,3,4\}$-逆) *Moore–Penrose* 广义逆 A^+ 是同时满足所有四个条件的广义逆：

1. $AA^+A=A$ （$\{1\}$）
2. $A^+AA^+=A^+$ （$\{2\}$）
3. $(AA^+)^*=AA^+$ （$\{3\}$）
4. $(A^+A)^*=A^+A$ （$\{4\}$）

因此，A^+ 是 $\{1,2,3,4\}$-逆。

- $\{1\}$-逆：一般不唯一（除非 A 可逆）
- $\{1,2\}$-逆：一般不唯一
- $\{1,3\}$-逆：唯一（若存在）
- $\{1,4\}$-逆：唯一（若存在）
- $\{1,2,3,4\}$-逆（M-P 逆）：唯一

12.6 Exercises

13 Linear Transformations VI 线性变换 VI（从对角矩阵推广至 JNF）

13.1 The Jordan Normal Form (JNF)

对于复方阵 $A \in \mathbb{C}^{n\times n}$，即使它不可对角化，我们仍然希望找到一个尽可能简单的相似标准形。Jordan 标准形（JNF）实现了这一目标：它是除排列外唯一的、最接近对角形的相似标准形。

定义 13.1 (Jordan 块) *一个 $k \times k$ 的* ***Jordan* 块** *对应特征值 λ 定义为：*

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}_{k\times k}.$$

即主对角元全为 λ，上次对角元全为 1。

定义 13.2 (Jordan 矩阵) *形如*

$$J = \begin{pmatrix} J_{k_1}(\lambda_1) & & & \\ & J_{k_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{k_m}(\lambda_m) \end{pmatrix}$$

的分块对角矩阵称为 ***Jordan* 矩阵**。*每个 $J_{k_i}(\lambda_i)$ 是一个 Jordan 块，λ_i 可以相同（对应同一特征值的不同块）。*

定理 13.1 (Jordan 标准形定理) *设 $A \in \mathbb{C}^{n\times n}$，则存在可逆矩阵 $P \in \mathbb{C}^{n\times n}$ 和 Jordan 矩阵 J，使得*

$$A = PJP^{-1}.$$

J 在 Jordan 块的排列次序下是唯一的。

在 A 的 JNF 中：

- 特征值 λ 对应的 Jordan 块个数 = 几何重数 g_λ
- 所有以 λ 为特征值的 Jordan 块的阶数之和 = 代数重数 a_λ

定义 13.3 (广义特征向量) *设 λ 是 A 的特征值，向量 $\mathbf{v} \neq \mathbf{0}$ 称为对应 λ 的* **级数为 m 的广义特征向量**，*若*

$$(A-\lambda I)^m \mathbf{v} = \mathbf{0} \quad \text{且} \quad (A-\lambda I)^{m-1}\mathbf{v} \neq \mathbf{0}.$$

设 $\mathbf{v}_m$ 是级数为 m 的广义特征向量，定义

$$\mathbf{v}_{m-1} = (A-\lambda I)\mathbf{v}_m, \quad \mathbf{v}_{m-2} = (A-\lambda I)\mathbf{v}_{m-1}, \quad \ldots, \quad \mathbf{v}_1 = (A-\lambda I)\mathbf{v}_2.$$

则 $\mathbf{v}_1$ 是普通特征向量，且 $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ 称为一个 **Jordan 链**。

在 JNF 中，P 的列由各个 Jordan 链的向量组成。这些向量构成 $\mathbb{C}^n$ 的一组基，称为 **Jordan 基**。

JNF 的计算步骤

步骤 1：求特征值与代数重数计算特征多项式 $p_A(\lambda) = \det(A - \lambda I)$，分解得特征值 $\lambda_1, \ldots, \lambda_k$ 及其代数重数 a_{λ_i}。

步骤 2：对每个特征值求广义特征空间对每个特征值 λ，计算：

$$\begin{aligned}
&E_1(\lambda) = \text{null}(A - \lambda I) \quad (\text{特征空间}) \\
&E_2(\lambda) = \text{null}((A - \lambda I)^2) \\
&\vdots \\
&E_m(\lambda) = \text{null}((A - \lambda I)^m)
\end{aligned}$$

直到 $E_m(\lambda) = E_{m+1}(\lambda)$（稳定）。

步骤 3：确定 Jordan 块结构对每个 λ：

1. 几何重数 $g_\lambda = \dim E_1(\lambda) =$ Jordan 块个数
2. 设 $d_i = \dim E_i(\lambda) - \dim E_{i-1}(\lambda)$（其中 $E_0(\lambda) = \{\mathbf{0}\}$）
3. 大小为 $\geq i$ 的 Jordan 块个数 $= d_i$

步骤 4：构造 Jordan 链和矩阵 P 对每个 Jordan 块 $J_k(\lambda)$，从链尾开始：

1. 选取 $\mathbf{v}_k \in E_k(\lambda) \setminus E_{k-1}(\lambda)$
2. 计算 $\mathbf{v}_{k-1} = (A - \lambda I)\mathbf{v}_k$
3. 继续直到 $\mathbf{v}_1 = (A - \lambda I)\mathbf{v}_2$（特征向量）
4. 这些向量构成 P 中对应该 Jordan 块的列

例题 13.1 *求* $A = \begin{pmatrix} 4 & -1 & 1 \\ 1 & 2 & 1 \\ -2 & 1 & 1 \end{pmatrix}$ *的 JNF。*

解：

1. 特征多项式：$p_A(\lambda) = (\lambda - 3)^3$，$\lambda = 3$ 为三重特征值，$a_3 = 3$。

2. 计算：

$$A - 3I = \begin{pmatrix} 1 & -1 & 1 \\ 1 & -1 & 1 \\ -2 & 1 & -2 \end{pmatrix}.$$

$\text{rank}(A - 3I) = 2 \Rightarrow \dim E_1(3) = 1$，*故* $g_3 = 1$。

3. 几何重数为 1，说明只有一个 Jordan 块。

4. 计算 $(A - 3I)^2 = 0$（幂零），故最大 Jordan 链长为 3。

5. 选取 $\mathbf{v}_3 \notin E_1(3)$，例如 $\mathbf{v}_3 = (0, 0, 1)^T$。

6. 计算 Jordan 链：

$$\mathbf{v}_2 = (A - 3I)\mathbf{v}_3 = (1, 1, -2)^T, \quad \mathbf{v}_1 = (A - 3I)\mathbf{v}_2 = (0, 0, 0)^T? \quad \text{检查！}$$

实际上：$\mathbf{v}_2$ 已在线性相关。重选 $\mathbf{v}_3 = (1,0,0)^T$：

$$\mathbf{v}_2 = (A-3I)\mathbf{v}_3 = (1,1,-2)^T, \quad \mathbf{v}_1 = (A-3I)\mathbf{v}_2 = (0,0,0)^T.$$

但 $(A-3I)^2\mathbf{v}_3 = 0$，说明链长为 *2*？我们需要重新考虑。

实际上，正确计算得：

$$P = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ -2 & 1 & 0 \end{pmatrix}, \quad J = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

即两个 *Jordan* 块：一个 2×2 块 $J_2(3)$ 和一个 1×1 块 $J_1(3)$。

2×2			
linearly independent eigenvector 的个数	eigenvalue	是否可被对角化 (LI eigenvector 的个数是否足够)	
2	2 个不同	✓	
2	2 个一样	✓	
1	2 个一样	× (JNF)	
$J = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$ $(A-\lambda I_2)v_2 = v_1$			
3×3			
linearly independent eigenvector 的个数	eigenvalue	是否可被对角化 (LI eigenvector 的个数是否足够)	
3	3 个不同	✓	
3	2 个一样	✓	
2	2 个一样	× (JNF)	case 1
3	3 个一样	✓	
2	3 个一样	× (JNF)	case 2
1	3 个一样	× (JNF)	case 3
case 1: $J = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \mu \end{pmatrix}$ $(A-\lambda I_3)v_2 = v_1$ case 2 (take care!): $J = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}$ $(A-\lambda I_3)v_2 = v_1$ v_3 to be an eigenvector for λ that is not a scalar multiple of v_1 case 3: $J = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$ $(A-\lambda I_3)v_2 = v_1$ $(A-\lambda I_3)v_3 = v_2$			

13.2 Application Scenario 1: Solving Systems of Difference Equations

13.3 Application Scenario 2: Solving Systems of Differential Equations

13.4 Application Scenario 3: Dominant Eigenvalues and Long-Term Behaviour

13.5 Exercises

14 Matrix Algebra 矩阵代数

14.1 Review: Basics

$$(AB)^{-1} = B^{-1}A^{-1} \tag{14.1}$$
$$(ABC\cdots)^{-1} = \cdots C^{-1}B^{-1}A^{-1} \tag{14.2}$$
$$(A^T)^{-1} = (A^{-1})^T \tag{14.3}$$
$$(A+B)^T = A^T + B^T \tag{14.4}$$
$$(AB)^T = B^T A^T \tag{14.5}$$
$$(ABC\cdots)^T = \cdots C^T B^T A^T \tag{14.6}$$
$$(A^H)^{-1} = (A^{-1})^H \tag{14.7}$$
$$(A+B)^H = A^H + B^H \tag{14.8}$$
$$(AB)^H = B^H A^H \tag{14.9}$$
$$(ABC\cdots)^H = \cdots C^H B^H A^H \tag{14.10}$$
$$\mathrm{Tr}(A) = \sum_i A_{ii} \tag{14.11}$$
$$\mathrm{Tr}(A) = \sum_i \lambda_i, \quad \lambda_i = \mathrm{eig}(A) \tag{14.12}$$
$$\mathrm{Tr}(A) = \mathrm{Tr}(A^T) \tag{14.13}$$
$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA) \tag{14.14}$$
$$\mathrm{Tr}(A+B) = \mathrm{Tr}(A) + \mathrm{Tr}(B) \tag{14.15}$$
$$\mathrm{Tr}(ABC) = \mathrm{Tr}(BCA) = \mathrm{Tr}(CAB) \tag{14.16}$$
$$\mathbf{a}^T\mathbf{a} = \mathrm{Tr}(\mathbf{a}\mathbf{a}^T) \tag{14.17}$$

A 是一个 $n \times n$ 矩阵。

$$\det(A) = \prod_i \lambda_i, \quad \lambda_i = \mathrm{eig}(A) \tag{14.18}$$
$$\det(cA) = c^n \det(A), \quad \text{若} A \in \mathbb{R}^{n\times n} \tag{14.19}$$
$$\det(A^T) = \det(A) \tag{14.20}$$
$$\det(AB) = \det(A)\det(B) \tag{14.21}$$
$$\det(A^{-1}) = 1/\det(A) \tag{14.22}$$
$$\det(A^n) = \det(A)^n \tag{14.23}$$
$$\det(I + uv^T) = 1 + u^T v \tag{14.24}$$

对于 $n = 2$:

$$\det(I + A) = 1 + \det(A) + \mathrm{Tr}(A) \tag{14.25}$$

对于 $n = 3$:

$$\det(I + A) = 1 + \det(A) + \mathrm{Tr}(A) + \frac{1}{2}\mathrm{Tr}(A)^2 - \frac{1}{2}\mathrm{Tr}(A^2) \tag{14.26}$$

对于小的 ε，有以下近似：

$$\det(I+\varepsilon A) \cong 1+\det(A)+\varepsilon \operatorname{Tr}(A)+\frac{1}{2}\varepsilon^2 \operatorname{Tr}(A)^2-\frac{1}{2}\varepsilon^2 \operatorname{Tr}(A^2) \tag{14.27}$$

14.2 Derivatives of a Determinant

14.2.1 General form

$$\frac{\partial \det(\mathbf{Y})}{\partial x} = \det(\mathbf{Y}) \operatorname{Tr}\left[\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\right] \tag{14.28}$$

$$\sum_k \frac{\partial \det(\mathbf{X})}{\partial X_{ik}} X_{jk} = \delta_{ij} \det(\mathbf{X}) \tag{14.29}$$

$$\frac{\partial^2 \det(\mathbf{Y})}{\partial x^2} = \det(\mathbf{Y}) \left[\operatorname{Tr}\left[\mathbf{Y}^{-1}\frac{\partial \frac{\partial \mathbf{Y}}{\partial x}}{\partial x}\right]\right. \tag{14.30}$$

$$+ \operatorname{Tr}\left[\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\right] \operatorname{Tr}\left[\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\right] \tag{14.31}$$

$$\left. - \operatorname{Tr}\left[\left(\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\right)\left(\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\right)\right]\right] \tag{14.32}$$

14.2.2 Linear forms

$$\frac{\partial \det(X)}{\partial X} = \det(X)(X^{-1})^T \tag{14.33}$$

$$\sum_k \frac{\partial \det(X)}{\partial X_{ik}} X_{jk} = \delta_{ij} \det(X) \tag{14.34}$$

$$\frac{\partial \det(AXB)}{\partial X} = \det(AXB)(X^{-1})^T = \det(AXB)(X^T)^{-1} \tag{14.35}$$

14.2.3 Square forms

若 X 为方阵且可逆，则

$$\frac{\partial \det(X^TAX)}{\partial X} = 2\det(X^TAX)X^{-T} \tag{14.36}$$

若 X 非方阵但 A 对称，则

$$\frac{\partial \det(X^TAX)}{\partial X} = 2\det(X^TAX)AX(X^TAX)^{-1} \tag{14.37}$$

若 X 非方阵且 A 不对称，则

$$\frac{\partial \det(X^TAX)}{\partial X} = \det(X^TAX)(AX(X^TAX)^{-1}+A^TX(X^TA^TX)^{-1}) \tag{14.38}$$

14.2.4 Other nonlinear forms

$$\frac{\partial \ln\det(X^TX)}{\partial X} = 2(X^+)^T \tag{14.39}$$

$$\frac{\partial \ln \det(X^T X)}{\partial X^+} = -2X^T \tag{14.40}$$

$$\frac{\partial \ln |\det(X)|}{\partial X} = (X^{-1})^T = (X^T)^{-1} \tag{14.41}$$

$$\frac{\partial \det(X^k)}{\partial X} = k \det(X^k) X^{-T} \tag{14.42}$$

14.3 Derivatives of an Inverse

$$\frac{\partial Y^{-1}}{\partial x} = -Y^{-1} \frac{\partial Y}{\partial x} Y^{-1} \tag{14.43}$$

$$\frac{\partial (X^{-1})_{kl}}{\partial X_{ij}} = -(X^{-1})_{ki} (X^{-1})_{jl} \tag{14.44}$$

$$\frac{\partial a^T X^{-1} b}{\partial X} = -X^{-T} a b^T X^{-T} \tag{14.45}$$

$$\frac{\partial \det(X^{-1})}{\partial X} = -\det(X^{-1}) (X^{-1})^T \tag{14.46}$$

$$\frac{\partial \operatorname{Tr}(AX^{-1}B)}{\partial X} = -(X^{-1}BAX^{-1})^T \tag{14.47}$$

$$\frac{\partial \operatorname{Tr}((X+A)^{-1})}{\partial X} = -((X+A)^{-1}(X+A)^{-1})^T \tag{14.48}$$

设 A 为 $n \times n$ 可逆方阵，W 是 A 的逆矩阵，$J(A)$ 是关于 A 的可微 $n \times n$ 矩阵函数，则 J 对 A 和 W 的偏微分满足

$$\frac{\partial J}{\partial A} = -A^{-T} \frac{\partial J}{\partial W} A^{-T} \tag{14.49}$$

14.4 Derivatives of Eigenvalues

$$\frac{\partial}{\partial X} \sum_{\operatorname{eig}(X)} \operatorname{eig}(X) = \frac{\partial}{\partial X} \operatorname{Tr}(X) = I \tag{14.50}$$

$$\frac{\partial}{\partial X} \prod_{\operatorname{eig}(X)} \operatorname{eig}(X) = \frac{\partial}{\partial X} \det(X) = \det(X) X^{-T} \tag{14.51}$$

若 A 是实对称矩阵，λ_i 和 v_i 是 A 的互异特征值和特征向量，且 $v_i^T v_i = 1$，则

$$\partial \lambda_i = v_i^T \partial(A) v_i \tag{14.52}$$

$$\partial v_i = (\lambda_i I - A)^+ \partial(A) v_i \tag{14.53}$$

14.5 Derivatives of Matrices, Vectors and Scalar Forms

14.5.1 First Order

$$\frac{\partial \mathbf{x}^T \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} \tag{14.54}$$

$$\frac{\partial \mathbf{a}^T X \mathbf{b}}{\partial X} = \mathbf{a}\mathbf{b}^T \tag{14.55}$$

$$\frac{\partial \mathbf{a}^T X^T \mathbf{b}}{\partial X} = \mathbf{b}\mathbf{a}^T \tag{14.56}$$

$$\frac{\partial \mathbf{a}^T X \mathbf{a}}{\partial X} = \frac{\partial \mathbf{a}^T X^T \mathbf{a}}{\partial X} = \mathbf{a}\mathbf{a}^T \tag{14.57}$$

$$\frac{\partial X}{\partial X_{ij}} = J^{ij} \tag{14.58}$$

$$\frac{\partial (XA)_{ij}}{\partial X_{mn}} = \delta_{im}(A)_{nj} = (J^{mn}A)_{ij} \tag{14.59}$$

$$\frac{\partial (X^T A)_{ij}}{\partial X_{mn}} = \delta_{in}(A)_{mj} = (J^{nm}A)_{ij} \tag{14.60}$$

14.5.2 Second Order

$$\frac{\partial}{\partial X_{ij}} \sum_{klmn} X_{kl} X_{mn} = 2 \sum_{kl} X_{kl} \tag{14.61}$$

$$\frac{\partial \mathbf{b}^T X^T X \mathbf{c}}{\partial X} = X(\mathbf{b}\mathbf{c}^T + \mathbf{c}\mathbf{b}^T) \tag{14.62}$$

$$\frac{\partial (B\mathbf{x} + \mathbf{b})^T C (D\mathbf{x} + \mathbf{d})}{\partial \mathbf{x}} = B^T C (D\mathbf{x} + \mathbf{d}) + D^T C^T (B\mathbf{x} + \mathbf{b}) \tag{14.63}$$

$$\frac{\partial (X^T B X)_{kl}}{\partial X_{ij}} = \delta_{ij}(X^T B)_{kl} + \delta_{kj}(BX)_{il} \tag{14.64}$$

$$\frac{\partial (X^T B X)}{\partial X_{ij}} = X^T B J^{ij} + J^{ji} B X \quad (J^{ij})_{kl} = \delta_{ik}\delta_{jl} \tag{14.65}$$

$$\frac{\partial \mathbf{x}^T B \mathbf{x}}{\partial \mathbf{x}} = (B + B^T)\mathbf{x} \tag{14.66}$$

$$\frac{\partial \mathbf{b}^T X^T D X \mathbf{c}}{\partial X} = D^T X \mathbf{b}\mathbf{c}^T + D X \mathbf{c}\mathbf{b}^T \tag{14.67}$$

$$\frac{\partial}{\partial X}(X\mathbf{b} + \mathbf{c})^T D (X\mathbf{b} + \mathbf{c}) = (D + D^T)(X\mathbf{b} + \mathbf{c})\mathbf{b}^T \tag{14.68}$$

假设 W 对称，则

$$\frac{\partial}{\partial \mathbf{s}}(\mathbf{x} - A\mathbf{s})^T W (\mathbf{x} - A\mathbf{s}) = -2A^T W (\mathbf{x} - A\mathbf{s}) \tag{14.69}$$

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x} - \mathbf{s})^T W (\mathbf{x} - \mathbf{s}) = 2W(\mathbf{x} - \mathbf{s}) \tag{14.70}$$

$$\frac{\partial}{\partial \mathbf{s}}(\mathbf{x} - \mathbf{s})^T W (\mathbf{x} - \mathbf{s}) = -2W(\mathbf{x} - \mathbf{s}) \tag{14.71}$$

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x} - A\mathbf{s})^T W (\mathbf{x} - A\mathbf{s}) = 2W(\mathbf{x} - A\mathbf{s}) \tag{14.72}$$

$$\frac{\partial}{\partial A}(\mathbf{x} - A\mathbf{s})^T W (\mathbf{x} - A\mathbf{s}) = -2W(\mathbf{x} - A\mathbf{s})\mathbf{s}^T \tag{14.73}$$

对于复数情形：

$$\frac{\partial (a - \mathbf{x}^H \mathbf{b})^2}{\partial \mathbf{x}} = -2\mathbf{b}(a - \mathbf{x}^H \mathbf{b})^* \tag{14.74}$$

14.5.3 Higher-order and non-linear

$$\frac{\partial (X^n)_{kl}}{\partial X_{ij}} = \sum_{r=0}^{n-1} (X^r J^{ij} X^{n-1-r})_{kl} \tag{14.75}$$

$$\frac{\partial}{\partial X}\mathbf{a}^T X^n \mathbf{b} = \sum_{r=0}^{n-1}(X^r)^T \mathbf{a}\mathbf{b}^T (X^{n-1-r})^T \tag{14.76}$$

$$\frac{\partial}{\partial X}\mathbf{a}^T (X^n)^T X^n \mathbf{b} = \sum_{r=0}^{n-1}\left[X^{n-1-r}\mathbf{a}\mathbf{b}^T (X^n)^T X^r + (X^r)^T X^n \mathbf{a}\mathbf{b}^T (X^{n-1-r})^T\right] \tag{14.77}$$

假设 $\mathbf{s}=\mathbf{s}(\mathbf{x}), \mathbf{r}=\mathbf{r}(\mathbf{x})$，且 A 为常数，则

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{s}^T A\mathbf{r} = \left[\frac{\partial \mathbf{s}}{\partial \mathbf{x}}\right]^T A\mathbf{r} + \left[\frac{\partial \mathbf{r}}{\partial \mathbf{x}}\right]^T A^T \mathbf{s} \tag{14.78}$$

14.5.4 Gradient and Hessian

$$f = \mathbf{x}^T A\mathbf{x} + \mathbf{b}^T \mathbf{x} \tag{14.79}$$

$$\nabla_{\mathbf{x}} f = \frac{\partial f}{\partial \mathbf{x}} = (A + A^T)\mathbf{x} + \mathbf{b} \tag{14.80}$$

$$\frac{\partial^2 f}{\partial \mathbf{x}\partial \mathbf{x}^T} = A + A^T \tag{14.81}$$

14.6 Derivatives of Traces

设 $F(X)$ 是 X 的每个元素的可微函数，则

$$\frac{\partial \operatorname{Tr}(F(X))}{\partial X} = f(X)^T$$

其中 $f(\cdot)$ 是 $F(\cdot)$ 的标量导数。

14.6.1 First Order

$$\frac{\partial}{\partial X}\operatorname{Tr}(X) = I \tag{14.82}$$

$$\frac{\partial}{\partial X}\operatorname{Tr}(XA) = A^T \tag{14.83}$$

$$\frac{\partial}{\partial X}\operatorname{Tr}(AXB) = A^T B^T \tag{14.84}$$

$$\frac{\partial}{\partial X}\operatorname{Tr}(AX^T B) = BA \tag{14.85}$$

$$\frac{\partial}{\partial X}\operatorname{Tr}(X^T A) = A \tag{14.86}$$

$$\frac{\partial}{\partial X}\operatorname{Tr}(AX^T) = A \tag{14.87}$$

$$\frac{\partial}{\partial X}\operatorname{Tr}(A \otimes X) = \operatorname{Tr}(A)I \tag{14.88}$$

14.6.2 Second Order

$$\frac{\partial}{\partial X}\operatorname{Tr}(X^2) = 2X^T \tag{14.89}$$

$$\frac{\partial}{\partial X}\operatorname{Tr}(X^2 B) = (XB + BX)^T \tag{14.90}$$

$$\frac{\partial}{\partial X}\operatorname{Tr}(X^TBX) = BX + B^TX \tag{14.91}$$
$$\frac{\partial}{\partial X}\operatorname{Tr}(BXX^T) = BX + B^TX \tag{14.92}$$
$$\frac{\partial}{\partial X}\operatorname{Tr}(XX^TB) = BX + B^TX \tag{14.93}$$
$$\frac{\partial}{\partial X}\operatorname{Tr}(XBX^T) = XB^T + XB \tag{14.94}$$
$$\frac{\partial}{\partial X}\operatorname{Tr}(BX^TX) = XB^T + XB \tag{14.95}$$
$$\frac{\partial}{\partial X}\operatorname{Tr}(X^TXB) = XB^T + XB \tag{14.96}$$
$$\frac{\partial}{\partial X}\operatorname{Tr}(AXBX) = A^TX^TB^T + B^TX^TA^T \tag{14.97}$$
$$\frac{\partial}{\partial X}\operatorname{Tr}(X^TX) = \frac{\partial}{\partial X}\operatorname{Tr}(XX^T) = 2X \tag{14.98}$$
$$\frac{\partial}{\partial X}\operatorname{Tr}(B^TX^TCXB) = C^TXBB^T + CXBB^T \tag{14.99}$$
$$\frac{\partial}{\partial X}\operatorname{Tr}(X^TBXC) = BXC + B^TXC^T \tag{14.100}$$
$$\frac{\partial}{\partial X}\operatorname{Tr}(AXBX^TC) = A^TC^TXB^T + CAXB \tag{14.101}$$
$$\frac{\partial}{\partial X}\operatorname{Tr}[(AXB+C)(AXB+C)^T] = 2A^T(AXB+C)B^T \tag{14.102}$$
$$\frac{\partial}{\partial X}\operatorname{Tr}(X\otimes X) = \frac{\partial}{\partial X}\operatorname{Tr}(X)\operatorname{Tr}(X) = 2\operatorname{Tr}(X)I \tag{14.103}$$

14.6.3 Higher Order

$$\frac{\partial}{\partial X}\operatorname{Tr}(X^k) = k(X^{k-1})^T \tag{14.104}$$
$$\frac{\partial}{\partial X}\operatorname{Tr}(AX^k) = \sum_{r=0}^{k-1}(X^rAX^{k-r-1})^T \tag{14.105}$$
$$\frac{\partial}{\partial X}\operatorname{Tr}[B^TX^TCXX^TCXB] = CXX^TCXBB^T + C^TXBB^TX^TC^TX \tag{14.106}$$
$$+ CXBB^TX^TCX + C^TXX^TC^TXBB^T \tag{14.107}$$

14.6.4 Other

$$\frac{\partial}{\partial X}\operatorname{Tr}(AX^{-1}B) = -(X^{-1}BAX^{-1})^T = -X^{-T}A^TB^TX^{-T} \tag{14.108}$$

假设 B 和 C 对称，则

$$\frac{\partial}{\partial X}\operatorname{Tr}[(X^TCX)^{-1}A] = -(CX(X^TCX)^{-1})(A+A^T)(X^TCX)^{-1} \tag{14.109}$$
$$\frac{\partial}{\partial X}\operatorname{Tr}[(X^TCX)^{-1}(X^TBX)] = -2CX(X^TCX)^{-1}X^TBX(X^TCX)^{-1} \tag{14.110}$$
$$+ 2BX(X^TCX)^{-1} \tag{14.111}$$
$$\frac{\partial}{\partial X}\operatorname{Tr}[(A+X^TCX)^{-1}(X^TBX)] = -2CX(A+X^TCX)^{-1}X^TBX(A+X^TCX)^{-1} \tag{14.112}$$

$$+2BX(A+X^TCX)^{-1} \tag{14.113}$$

$$\frac{\partial\,\mathrm{Tr}(\sin(X))}{\partial X} = \cos(X)^T \tag{14.114}$$

14.7 Derivatives of vector norms

$$\frac{\partial}{\partial \mathbf{x}}\|\mathbf{x}-\mathbf{a}\|_2 = \frac{\mathbf{x}-\mathbf{a}}{\|\mathbf{x}-\mathbf{a}\|_2} \tag{14.115}$$

$$\frac{\partial}{\partial \mathbf{x}}\frac{\mathbf{x}-\mathbf{a}}{\|\mathbf{x}-\mathbf{a}\|_2} = \frac{I}{\|\mathbf{x}-\mathbf{a}\|_2} - \frac{(\mathbf{x}-\mathbf{a})(\mathbf{x}-\mathbf{a})^T}{\|\mathbf{x}-\mathbf{a}\|_2^3} \tag{14.116}$$

$$\frac{\partial\|\mathbf{x}\|_2^2}{\partial \mathbf{x}} = \frac{\partial\|\mathbf{x}^T\mathbf{x}\|_2}{\partial \mathbf{x}} = 2\mathbf{x} \tag{14.117}$$

14.8 Matrix Decompositions

14.8.1 LU Decomposition

- **适用条件**:
  - 充分条件：矩阵 A 的所有顺序主子式均不为零。这意味着矩阵 A 的 1×1 左上角子矩阵、2×2 左上角子矩阵、……、直到 $n\times n$ 整个矩阵的行列式都不为零。在此条件下，可以不进行行交换（即不使用置换矩阵 P）直接进行高斯消元，得到唯一的 LU 分解（L 是单位下三角矩阵，U 是上三角矩阵）。
  - 必要条件：矩阵 A 必须是非奇异的（即可逆的）。如果矩阵是奇异的（行列式为零），则即使进行行交换，也无法得到具有非零对角元的 U 矩阵。不过，对于奇异矩阵，有时可以分解为 PLU 形式，但 U 的最后一（或多）行全为零。
- **形式**: $A=LU$，其中 L 为单位下三角矩阵，U 为上三角矩阵
  - 在实际数值计算中，即使顺序主子式不为零，如果出现很小的主元，也会导致数值不稳定。因此，通常使用带部分主元选取的 PLU 分解（即 $PA=LU$，其中 P 是置换矩阵）。这种分解对任何可逆矩阵 A 都存在。
  - 对称形式: $A=LDU$
- **几何意义**: 高斯消元法的矩阵表示，将线性变换分解为顺序行操作
- **主要用途**:
  - 解线性方程组 $A\mathbf{x}=\mathbf{b}$
  - 计算行列式 $\det(A)=\prod u_{ii}$
  - 求逆矩阵
  - 数值计算基础

条件/情形	能否 LU 分解	是否需要行交换	备注
所有顺序主子式 $\neq 0$	**是**，且唯一	不需要（$A = LU$）	理想情况，可进行朴素高斯消元
A **可逆，但顺序主子式不满足**	**是**，但不唯一	需要（$PA = LU$）	最常见的情况，使用部分主元法
A **奇异（不可逆）**	**是**，但不唯一	可能需要 P	U 的最后一（多）行全为零
A **是方阵且正定/对称正定**	**是**，且唯一	不需要	可进行**楚列斯基分解**（Cholesky），即 $A = LL^T$

14.8.2 QR Decomposition

- **适用条件**: 任意 $m \times n$ 矩阵（实或复）
- **形式**: $A = QR$
  - Q: $m \times m$ 正交（酉）矩阵，$Q^TQ = I$
  - R: $m \times n$ 上三角矩阵
- **几何意义**: Gram-Schmidt 正交化过程的矩阵表示，将列空间正交化
- **主要用途**:
  - 解最小二乘问题 $\min \|A\mathbf{x} - \mathbf{b}\|$
  - 计算特征值（QR 算法）
  - 正交基构造
  - 数值稳定求解线性系统

14.8.3 Eigen Decomposition (Diagonalisation)

- **适用条件**: 可对角化方阵（有 n 个线性无关特征向量）
- **形式**: $A = PDP^{-1}$
  - P: 特征向量矩阵（列向量为特征向量）
  - D: 对角矩阵，对角线为特征值 λ_i
- **几何意义**: 将变换分解为沿特征向量方向的伸缩变换
- **主要用途**:
  - 计算矩阵幂 $A^k = PD^kP^{-1}$
  - 分析动力系统 $\mathbf{x}_{k+1} = A\mathbf{x}_k$
  - 解微分方程组
  - 主成分分析（PCA）基础

14.8.4 Singular Value Decomposition

- **适用条件**: 任意 $m \times n$ 矩阵
- **形式**: $A = U\Sigma V^T$
  - U: $m \times m$ 正交矩阵（左奇异向量）
  - Σ: $m \times n$ 对角矩阵，对角线为奇异值 $\sigma_i \geq 0$
  - V: $n \times n$ 正交矩阵（右奇异向量）
- **几何意义**: 任何线性变换可分解为旋转/反射-伸缩-旋转/反射
- **主要用途**:
  - 矩阵低秩近似（图像压缩）
  - 解病态最小二乘问题
  - 计算矩阵秩、范数、条件数
  - 推荐系统、自然语言处理
  - 广义逆计算 $A^+ = V\Sigma^+U^T$

14.8.5 Cholesky Decomposition

- **适用条件**: 对称正定矩阵
- **形式**: $A = LL^T$（或 $A = R^TR$）
  - L: 下三角矩阵，对角线元素为正
- **几何意义**: 对称正定矩阵的" 平方根" 分解
- **主要用途**:
  - 解对称正定线性方程组（比 LU 快约 2 倍）
  - 蒙特卡洛模拟中生成相关随机变量
  - 卡尔曼滤波中的协方差更新
  - 优化问题的牛顿法

14.8.6 Schur Decomposition

- **适用条件**: 任意方阵
- **形式**: $A = QTQ^T$（实）或 $A = QTQ^*$（复）
  - Q: 正交（酉）矩阵
  - T: 上三角矩阵（实数为拟上三角，含 2×2 块）
- **几何意义**: 任何方阵都酉相似于上三角矩阵
- **主要用途**:

- 数值计算特征值（QR 算法的基础）
- 分析不可对角化矩阵
- 计算矩阵函数 $f(A)$
- 理论分析的标准形式

14.8.7 Spectral Decomposition

- **适用条件**: 正规矩阵（$AA^* = A^*A$）
- **形式**: $A = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^*$
  - λ_i: 特征值
  - $\mathbf{u}_i$: 标准正交特征向量
- **几何意义**: 沿正交方向的特征伸缩
- **主要用途**:
  - 厄米特矩阵（实对称）分析
  - 酉矩阵分析
  - 量子力学中的可观测量表示
  - 正交投影的线性组合

14.8.8 Polar Decomposition

- **适用条件**: 任意方阵
- **形式**: $A = UP$（右极分解）或 $A = P'U'$（左极分解）
  - U: 酉（正交）矩阵
  - P: 半正定厄米特矩阵，$P = (A^*A)^{1/2}$
- **几何意义**: 任何线性变换 = 旋转/反射 + 伸缩（顺序可交换）
- **主要用途**:
  - 连续介质力学（变形梯度分解）
  - 矩阵的最优逼近
  - 计算 SVD 的中间步骤
  - 量子计算中的门分解

14.9 Exercises

15 Applications 应用

15.1 Applications in Statistics

15.1.1 Covariance Matrix and Correlation Matrix

假设我们有一个 $n \times p$ 的数据矩阵 $\mathbf{X}$，其中 n 是样本数量，p 是变量数量：

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

每一列对应一个随机变量，每一行对应一个观测样本。

定义 15.1 (样本均值向量) 设 $\mathbf{1}_n = (1, 1, \ldots, 1)^\top \in \mathbb{R}^n$，则样本均值向量为：

$$\bar{\mathbf{x}} = \frac{1}{n}\mathbf{X}^\top \mathbf{1}_n = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix}$$

其中 $\bar{x}_j = \frac{1}{n}\sum_{i=1}^n x_{ij}$ 是第 j 个变量的样本均值。

定义 15.2 (中心化数据矩阵) 中心化数据矩阵 $\mathbf{X}_c$ 定义为：

$$\mathbf{X}_c = \mathbf{X} - \mathbf{1}_n \bar{\mathbf{x}}^\top = \begin{pmatrix} x_{11} - \bar{x}_1 & \cdots & x_{1p} - \bar{x}_p \\ \vdots & \ddots & \vdots \\ x_{n1} - \bar{x}_1 & \cdots & x_{np} - \bar{x}_p \end{pmatrix}$$

或者等价地，$\mathbf{X}_c = \mathbf{H}\mathbf{X}$，其中 $\mathbf{H} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top$ 是中心化矩阵（投影矩阵）。

定义 15.3 (样本协方差矩阵) 样本协方差矩阵 $\mathbf{S} \in \mathbb{R}^{p\times p}$ 定义为：

$$\mathbf{S} = \frac{1}{n-1}\mathbf{X}_c^\top \mathbf{X}_c = \frac{1}{n-1}\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$$

其中 $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ 是第 i 个样本。矩阵 $\mathbf{S}$ 的第 (j, k) 个元素为：

$$s_{jk} = \frac{1}{n-1}\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

特别地，对角线元素 $s_{jj} = s_j^2$ 是第 j 个变量的样本方差。

例题 15.1 设 $n = 3, p = 2$，数据矩阵为：

$$\mathbf{X} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{pmatrix}$$

则 $\bar{\mathbf{x}} = (2, 4)^\top$，中心化矩阵为：

$$\mathbf{X}_c = \begin{pmatrix} -1 & -2 \\ 0 & 0 \\ 1 & 2 \end{pmatrix}$$

协方差矩阵为：

$$\mathbf{S} = \frac{1}{2}\begin{pmatrix} (-1)^2 + 0^2 + 1^2 & (-1)(-2) + 0 \cdot 0 + 1 \cdot 2 \\ (-2)(-1) + 0 \cdot 0 + 2 \cdot 1 & (-2)^2 + 0^2 + 2^2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$$

命题 15.1 (协方差矩阵的基本性质) 设 $\mathbf{S}$ 为样本协方差矩阵，则：

1. $\mathbf{S}$ 是对称矩阵，即 $\mathbf{S}^\top = \mathbf{S}$
2. $\mathbf{S}$ 是半正定矩阵
3. $\mathrm{tr}(\mathbf{S}) = \sum_{j=1}^{p} s_j^2$，即所有变量的方差之和
4. $\mathbf{S}$ 的特征值非负

证明 15.1

(1) 对称性：$s_{jk} = s_{kj}$ 由定义直接可得。

(2) 半正定性：对任意 $\mathbf{v} \in \mathbb{R}^p$，

$$\mathbf{v}^\top \mathbf{S}\mathbf{v} = \frac{1}{n-1}\sum_{i=1}^{n} \mathbf{v}^\top (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top \mathbf{v} = \frac{1}{n-1}\sum_{i=1}^{n} [(\mathbf{x}_i - \bar{\mathbf{x}})^\top \mathbf{v}]^2 \geq 0$$

(3) 迹的性质：$\mathrm{tr}(\mathbf{S}) = \sum_{j=1}^{p} s_{jj} = \sum_{j=1}^{p} s_j^2$

(4) 由半正定性立即得到特征值非负。

定理 15.1 (协方差矩阵的谱分解) 设 $\mathbf{S}$ 的谱分解为：

$$\mathbf{S} = \mathbf{U} \quad \mathbf{U}^\top$$

其中 $\mathbf{U}$ 是正交矩阵， $= \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$，$\lambda_1 \geq \cdots \geq \lambda_p \geq 0$ 是 $\mathbf{S}$ 的特征值。这个分解对应主成分分析 *(PCA)*。

证明 15.2 由对称矩阵的谱分解定理，任意实对称矩阵可正交对角化。结合 $\mathbf{S}$ 半正定，特征值非负。

定义 15.4 (样本相关性矩阵) 设 $s_j = \sqrt{s_{jj}}$ 为第 j 个变量的样本标准差。样本相关性矩阵 $\mathbf{R} \in \mathbb{R}^{p\times p}$ 定义为：

$$\mathbf{R} = \mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2}$$

其中 $\mathbf{D} = \mathrm{diag}(s_1^2, s_2^2, \ldots, s_p^2)$。$\mathbf{R}$ 的第 (j, k) 个元素为：

$$r_{jk} = \frac{s_{jk}}{s_j s_k} = \frac{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2 \sum_{i=1}^{n}(x_{ik} - \bar{x}_k)^2}}$$

即 *Pearson* 相关系数。

例题 15.2 续前例，$\mathbf{S} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$，则：

$$\mathbf{D} = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}, \quad \mathbf{D}^{-1/2} = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}$$

$$\mathbf{R} = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

两个变量完全正相关。

命题 15.2 (相关性矩阵的性质) 设 $\mathbf{R}$ 为样本相关性矩阵，则：

1. $\mathbf{R}$ 是对称矩阵

2. $\mathbf{R}$ 是半正定矩阵

3. $r_{jj} = 1$，对所有 $j = 1, \ldots, p$

4. $|r_{jk}| \leq 1$，对所有 j, k

证明 15.3

(1) 由 $\mathbf{R}^\top = (\mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2})^\top = \mathbf{D}^{-1/2}\mathbf{S}^\top\mathbf{D}^{-1/2} = \mathbf{R}$。

(2) 对任意 $\mathbf{v} \in \mathbb{R}^p$，令 $\mathbf{w} = \mathbf{D}^{-1/2}\mathbf{v}$，则：

$$\mathbf{v}^\top\mathbf{R}\mathbf{v} = \mathbf{w}^\top\mathbf{S}\mathbf{w} \geq 0$$

因为 $\mathbf{S}$ 半正定。

(3) $r_{jj} = s_{jj}/(s_j s_j) = s_j^2/s_j^2 = 1$。

(4) 由 *Cauchy-Schwarz* 不等式：

$$|s_{jk}| = \left|\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)\right| \leq \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{ik} - \bar{x}_k)^2} = s_j s_k$$

故 $|r_{jk}| = |s_{jk}|/(s_j s_k) \leq 1$。

定理 15.2 (协方差矩阵的矩阵表示) 设 $\mathbf{X}_c$ 为中心化数据矩阵，则：

$$\mathbf{S} = \frac{1}{n-1}\mathbf{X}_c^\top\mathbf{X}_c$$

特别地，如果数据已经标准化（每个变量均值为 *0*，标准差为 *1*），即 $\mathbf{Z}$ 是标准化数据矩阵，则：

$$\mathbf{R} = \frac{1}{n-1}\mathbf{Z}^\top\mathbf{Z}$$

其中 $z_{ij} = (x_{ij} - \bar{x}_j)/s_j$。

证明 15.4 标准化数据的第 j 列为 $(x_{1j} - \bar{x}_j)/s_j, \ldots, (x_{nj} - \bar{x}_j)/s_j$，计算：

$$\frac{1}{n-1}\sum_{i=1}^{n} z_{ij}z_{ik} = \frac{1}{n-1}\sum_{i=1}^{n}\frac{x_{ij} - \bar{x}_j}{s_j}\frac{x_{ik} - \bar{x}_k}{s_k} = \frac{s_{jk}}{s_j s_k} = r_{jk}$$

即 $\frac{1}{n-1}\mathbf{Z}^\top\mathbf{Z} = \mathbf{R}$。

命题 15.3 (几何解释) 在多元统计中，$\mathbf{S}$ 可以看作数据点云的形状矩阵。数据点 $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^p$ 的样本协方差矩阵 $\mathbf{S}$ 的特征向量定义了数据点云的主轴方向，对应的特征值表示沿这些方向的方差大小。

证明 15.5 考虑数据点在方向 $\mathbf{u}$（单位向量）上的投影：$p_i = \mathbf{u}^\top\mathbf{x}_i$。投影的样本方差为：

$$\mathrm{Var}(\{p_i\}) = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{u}^\top\mathbf{x}_i - \mathbf{u}^\top\bar{\mathbf{x}})^2 = \mathbf{u}^\top\mathbf{S}\mathbf{u}$$

最大化 $\mathbf{u}^\top\mathbf{S}\mathbf{u}$ 在 $\|\mathbf{u}\| = 1$ 约束下的解是 $\mathbf{S}$ 的最大特征值对应的特征向量。

定理 15.3 (线性变换下的协方差矩阵) 设 $\mathbf{Y} = \mathbf{XA} + \mathbf{1}_n\mathbf{b}^\top$，其中 $\mathbf{A} \in \mathbb{R}^{p\times q}$，$\mathbf{b} \in \mathbb{R}^q$，则 $\mathbf{Y}$ 的协方差矩阵为：

$$\mathbf{S}_Y = \mathbf{A}^\top\mathbf{S}_X\mathbf{A}$$

特别地，如果 $\mathbf{A}$ 是非奇异方阵，则 $\mathbf{S}_X$ 和 $\mathbf{S}_Y$ 有相同的惯性（正特征值个数、零特征值个数、负特征值个数）。

证明 15.6 $\mathbf{Y}$ 的样本均值为 $\bar{\mathbf{y}} = \mathbf{A}^\top\bar{\mathbf{x}} + \mathbf{b}$，中心化矩阵为：

$$\mathbf{Y}_c = (\mathbf{XA} + \mathbf{1}_n\mathbf{b}^\top) - \mathbf{1}_n(\mathbf{A}^\top\bar{\mathbf{x}} + \mathbf{b})^\top = \mathbf{X}_c\mathbf{A}$$

因此：

$$\mathbf{S}_Y = \frac{1}{n-1}\mathbf{Y}_c^\top\mathbf{Y}_c = \frac{1}{n-1}\mathbf{A}^\top\mathbf{X}_c^\top\mathbf{X}_c\mathbf{A} = \mathbf{A}^\top\mathbf{S}_X\mathbf{A}$$

例题 15.3 (主成分分析) 主成分分析 *(PCA)* 中，我们寻找正交变换 $\mathbf{P}$ 使得 $\mathbf{Y} = \mathbf{X}_c\mathbf{P}$ 的协方差矩阵 $\mathbf{S}_Y = \mathbf{P}^\top\mathbf{SP}$ 为对角矩阵。由谱分解定理，取 $\mathbf{P} = \mathbf{U}$，则 $\mathbf{S}_Y =$ 。

定理 15.4 (相关性矩阵与标准化) 设 $\mathbf{Z}$ 为标准化数据矩阵，则：

1. $\bar{\mathbf{z}} = \mathbf{0}$（零向量）
2. $\mathbf{S}_Z = \mathbf{R}$
3. $\mathrm{tr}(\mathbf{R}) = p$

证明 15.7

(1) 标准化后每个变量均值为 *0*。

(2) 由前面定理已证。

(3) $\mathrm{tr}(\mathbf{R}) = \sum_{j=1}^p r_{jj} = \sum_{j=1}^p 1 = p$。

15.1.2 Principal Component Analysis (PCA)

主成分分析（PCA）是一种经典的降维技术，其核心思想是通过正交变换将一组可能相关的变量转换为一组线性不相关的变量，称为主成分。

定义 15.5 (PCA 的两种等价表述) 给定中心化数据矩阵 $\mathbf{X}_c \in \mathbb{R}^{n\times p}$（$n$ 个样本，p 个变量，已中心化），*PCA* 可以等价地表述为以下两个优化问题：

1. **最大方差观点**：寻找单位向量 $\mathbf{u}_1$，使得投影后的样本方差最大：

$$\max_{\|\mathbf{u}\|=1} \mathbf{u}^\top\mathbf{S}\mathbf{u}$$

其中 $\mathbf{S} = \frac{1}{n-1}\mathbf{X}_c^\top\mathbf{X}_c$ 是样本协方差矩阵。

2. **最小重构误差观点**：寻找 k 维子空间（$k < p$），使得数据点到该子空间投影的平方距离最小：

$$\min_{\mathbf{U}_k\in\mathbb{R}^{p\times k}, \mathbf{U}_k^\top\mathbf{U}_k=\mathbf{I}_k} \|\mathbf{X}_c - \mathbf{X}_c\mathbf{U}_k\mathbf{U}_k^\top\|_F^2$$

定理 15.5 (谱分解定理与 PCA) 设 $\mathbf{S} \in \mathbb{R}^{p\times p}$ 是对称半正定矩阵（协方差矩阵），则存在正交矩阵 $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_p]$ 和对角矩阵 $\ = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$，满足：

$$\mathbf{S} = \mathbf{U}\ \mathbf{U}^\top = \sum_{i=1}^p \lambda_i\mathbf{u}_i\mathbf{u}_i^\top$$

其中 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ 是 $\mathbf{S}$ 的特征值，$\mathbf{u}_i$ 是对应的单位特征向量。

定理 15.6 (PCA 的最优解) 对于第一主成分方向 $\mathbf{u}_1$ 的优化问题：

$$\max_{\|\mathbf{u}\|=1} \mathbf{u}^\top \mathbf{S}\mathbf{u}$$

最优解是 $\mathbf{S}$ 的最大特征值 λ_1 对应的特征向量 $\mathbf{u}_1$，且最大值为 λ_1。

更一般地，前 k 个主成分方向 $\mathbf{u}_1, \ldots, \mathbf{u}_k$ 是 $\mathbf{S}$ 的前 k 个最大特征值对应的特征向量，它们张成了最优的 k 维子空间。

PCA 算法步骤：

1. **数据预处理**：
   - 计算每列的均值：$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$
   - 中心化：$x_{ij}^{(c)} = x_{ij} - \bar{x}_j$，得到 $\mathbf{X}_c$
   - （可选）标准化：$z_{ij} = \frac{x_{ij}^{(c)}}{s_j}$，其中 s_j 是第 j 列的标准差
2. **计算协方差/相关矩阵**：
   - 若使用协方差矩阵：$\mathbf{S} = \frac{1}{n-1}\mathbf{X}_c^\top \mathbf{X}_c$
   - 若使用相关矩阵：$\mathbf{R} = \frac{1}{n-1}\mathbf{Z}^\top \mathbf{Z}$（标准化数据）
3. **特征值分解**：
   - 计算 $\mathbf{S}$（或 $\mathbf{R}$）的特征值和特征向量：$\mathbf{S} = \mathbf{U}\ \mathbf{U}^\top$
   - 特征值排序：$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$
   - 对应特征向量：$\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p$
4. **选择主成分数量**：
   - 计算累计方差贡献率：$\text{Cumulative}_k = \frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{p}\lambda_i}$
   - 选择最小的 k 使得 $\text{Cumulative}_k \geq \tau$（通常 $\tau = 0.8, 0.9, 0.95$）
5. **计算主成分得分**：
   - 载荷矩阵：$\mathbf{P} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k]$
   - 主成分得分矩阵：$\mathbf{T} = \mathbf{X}_c\mathbf{P}$
6. **输出**：$\mathbf{T}, \mathbf{P}, \{\lambda_i\}, \{\mathbf{u}_i\}$

定义 15.6 (方差解释) 第 i 个主成分解释的方差比例为：

$$\textit{Variance Explained}_i = \frac{\lambda_i}{\sum_{j=1}^{p}\lambda_j} = \frac{\lambda_i}{\text{tr}(\mathbf{S})}$$

前 k 个主成分累计解释的方差比例为：

$$\textit{Cumulative Variance}_k = \frac{\sum_{i=1}^{k}\lambda_i}{\sum_{j=1}^{p}\lambda_j}$$

例题 15.4 (方差解释计算) 设 $\mathbf{S}$ 的特征值为 $\lambda_1 = 5.2, \lambda_2 = 2.1, \lambda_3 = 0.8, \lambda_4 = 0.4$，则：

- 总方差：$\sum \lambda_i = 8.5$

- *PC1* 解释方差比例：$5.2/8.5 = 61.2\%$
- *PC2* 解释方差比例：$2.1/8.5 = 24.7\%$
- *PC1+PC2* 累计：$(5.2 + 2.1)/8.5 = 85.9\%$

命题 15.4 (主成分选择的准则) 常用准则包括：

1. **累计方差比例准则**：选择最小的 k 使得累计方差比例达到阈值（如 80%, 90%, 95%）
2. ***Kaiser* 准则**：保留特征值大于 1 的主成分（适用于相关矩阵）
3. **碎石图准则**：观察特征值下降的”肘部”点
4. **平行分析**：与随机数据特征值比较

定理 15.7 (主成分的正交性) 设 $\mathbf{P}_k = [\mathbf{u}_1, \ldots, \mathbf{u}_k]$ 是前 k 个主成分方向的矩阵，则：

1. $\mathbf{P}_k^\top \mathbf{P}_k = \mathbf{I}_k$（列正交）
2. 主成分得分 $\mathbf{T} = \mathbf{X}_c \mathbf{P}_k$ 的协方差矩阵为：

$$\mathrm{Cov}(\mathbf{T}) = \frac{1}{n-1} \mathbf{T}^\top \mathbf{T} = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$$

   即各主成分之间不相关，第 i 个主成分的方差为 λ_i

定理 15.8 (数据重构公式) 原始中心化数据可以通过主成分完美重构：

$$\mathbf{X}_c = \mathbf{T}\mathbf{P}^\top = \sum_{i=1}^{p} \mathbf{t}_i \mathbf{u}_i^\top$$

其中 $\mathbf{t}_i = \mathbf{X}_c \mathbf{u}_i$ 是第 i 个主成分得分向量。

使用前 k 个主成分的近似重构为：

$$\hat{\mathbf{X}}_c^{(k)} = \sum_{i=1}^{k} \mathbf{t}_i \mathbf{u}_i^\top = \mathbf{X}_c \mathbf{P}_k \mathbf{P}_k^\top$$

命题 15.5 (重构误差) 使用前 k 个主成分的重构误差（*Frobenius* 范数）为：

$$\|\mathbf{X}_c - \hat{\mathbf{X}}_c^{(k)}\|_F^2 = \sum_{i=k+1}^{p} \lambda_i$$

相对重构误差为：

$$\frac{\|\mathbf{X}_c - \hat{\mathbf{X}}_c^{(k)}\|_F^2}{\|\mathbf{X}_c\|_F^2} = \frac{\sum_{i=k+1}^{p} \lambda_i}{\sum_{i=1}^{p} \lambda_i}$$

定理 15.9 (PCA 的奇异值分解表示) 设 $\mathbf{X}_c$ 的奇异值分解（*SVD*）为：

$$\mathbf{X}_c = \mathbf{U} \ \mathbf{V}^\top$$

其中 $\mathbf{U} \in \mathbb{R}^{n \times p}$， $= \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$，$\mathbf{V} \in \mathbb{R}^{p \times p}$，且 $\sigma_1 \geq \cdots \geq \sigma_p \geq 0$。

则 *PCA* 结果可直接从 *SVD* 得到：

1. 主成分载荷（特征向量）：$\mathbf{V}$ 的列向量
2. 主成分得分：$\mathbf{T} = \mathbf{U}$

3. 特征值：$\lambda_i = \frac{\sigma_i^2}{n-1}$

注 15.1 (SVD 方法的优势) 在实践中，通常使用 *SVD* 而不是直接的特征值分解来计算 *PCA*，因为：

1. *SVD* 数值稳定性更好
2. 可直接处理 $n \ll p$ 的情况（高维小样本）
3. 不需要显式计算协方差矩阵 $\mathbf{S}$，节省内存

定义 15.7 (两种标准化方式) *PCA* 可以基于两种矩阵进行：

1. **协方差矩阵 *PCA***：使用 $\mathbf{S} = \frac{1}{n-1}\mathbf{X}_c^\top \mathbf{X}_c$
2. **相关矩阵 *PCA***：使用 $\mathbf{R} = \frac{1}{n-1}\mathbf{Z}^\top \mathbf{Z}$，其中 $\mathbf{Z}$ 是标准化数据

例题 15.5 (两种 PCA 的比较) 假设两个变量：X_1（身高，单位：米），X_2（体重，单位：千克）：

- 协方差矩阵 *PCA*：受量纲影响，方差大的变量（体重）主导第一主成分
- 相关矩阵 *PCA*：消除量纲影响，平等对待所有变量

命题 15.6 (选择准则)

- 使用协方差矩阵 *PCA*：当变量单位相同或量级相当时
- 使用相关矩阵 *PCA*：当变量单位不同或方差差异大时
- 相关矩阵 *PCA* = 标准化数据 + 协方差矩阵 *PCA*

定理 15.10 (主成分的抽样分布) 在大样本下（$n \to \infty$），若数据来自多元正态分布 $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，则：

1. 样本特征值 $\hat{\lambda}_i$ 是总体特征值 λ_i 的一致估计
2. 样本特征向量 $\hat{\mathbf{u}}_i$ 是总体特征向量 $\mathbf{u}_i$ 的一致估计
3. 当 λ_i 互不相同时，$\sqrt{n}(\hat{\mathbf{u}}_i - \mathbf{u}_i) \xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Gamma}_i)$

其中 $\boldsymbol{\Gamma}_i$ 的表达式复杂，依赖于 $\boldsymbol{\Sigma}$ 的特征值和特征向量。

定理 15.11 (特征值的渐近分布) 若 $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$，则：

$$\sqrt{n}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \xrightarrow{d} N_p(\mathbf{0}, 2\boldsymbol{\Lambda}^2)$$

即 $\hat{\lambda}_i$ 近似独立，且 $\mathrm{Var}(\hat{\lambda}_i) \approx \frac{2\lambda_i^2}{n}$。

例题 15.6 (PCA 的实际应用)

1. **数据可视化**：将高维数据降至 *2D* 或 *3D* 进行可视化
2. **噪声过滤**：保留主要成分，舍弃小特征值对应的成分
3. **特征提取**：在图像处理、信号处理中提取主要特征
4. **多重共线性检测**：小特征值表示变量间近似线性相关

定义 15.8 (核 PCA) 对于非线性数据，可使用核技巧：

1. 通过核函数 $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ 隐式映射到高维特征空间
2. 在高维空间中进行 *PCA*
3. 无需显式计算高维特征，只需核矩阵的特征值分解

定义 15.9 (稀疏 PCA) 通过添加 L_1 惩罚项获得稀疏载荷：

$$\max_{\|\mathbf{u}\|=1} \mathbf{u}^\top \mathbf{S}\mathbf{u} - \rho\|\mathbf{u}\|_1$$

使得主成分载荷中有许多零元素，提高可解释性。

例题 15.7 (Iris 数据集 PCA) 对著名的 *Iris* 数据集（$n = 150, p = 4$：萼片长宽、花瓣长宽）进行 *PCA*：

- 特征值：$\lambda_1 = 4.228, \lambda_2 = 0.243, \lambda_3 = 0.078, \lambda_4 = 0.024$
- 方差解释：*PC1* 解释 92.5%，*PC1+PC2* 解释 97.8%
- 载荷矩阵：

$$\mathbf{P} = \begin{pmatrix} 0.361 & -0.657 \\ 0.085 & -0.730 \\ 0.856 & 0.173 \\ 0.358 & 0.075 \end{pmatrix}$$

第一主成分主要反映花瓣长度，第二主成分反映萼片尺寸

定理 15.12 (PCA 与线性回归的关系) 考虑将 $\mathbf{y} \in \mathbb{R}^n$ 对 $\mathbf{X} \in \mathbb{R}^{n\times p}$ 的回归。若先对 $\mathbf{X}$ 进行 *PCA*，保留前 k 个主成分 $\mathbf{T}_k$，然后用 $\mathbf{y}$ 对 $\mathbf{T}_k$ 回归：

$$\hat{\mathbf{y}} = \mathbf{T}_k(\mathbf{T}_k^\top \mathbf{T}_k)^{-1}\mathbf{T}_k^\top \mathbf{y}$$

这就是主成分回归（*PCR*），可用于处理多重共线性。

定理 15.13 (PCA 主要性质总结) 设 $\mathbf{X}_c$ 为中心化数据矩阵，$\mathbf{S}$ 为其协方差矩阵，$\mathbf{S} = \mathbf{U}\ \mathbf{U}^\top$ 为谱分解，则：

1. **最优性**：前 k 个主成分张成的子空间是 k 维线性子空间中对数据方差保持最好的
2. **不相关性**：不同主成分之间互不相关
3. **方差最大化**：每个主成分最大化剩余方差
4. **嵌套性**：k 维 *PCA* 解包含在 $(k+1)$ 维解中
5. **旋转不变性**：*PCA* 结果在正交变换下具有不变性

15.2 Applications in Machine Learning

15.2.1 Sparse Matrix

定义 15.10 (稀疏矩阵) 设矩阵 $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{m\times n}$。如果 $\mathbf{A}$ 中非零元素的数量远少于零元素的数量，则称 $\mathbf{A}$ 为稀疏矩阵。形式上，定义稀疏度（*sparsity*）为：

$$Sparsity(\mathbf{A}) = 1 - \frac{nnz(\mathbf{A})}{mn}$$

其中 $nnz(\mathbf{A})$ 表示 $\mathbf{A}$ 中非零元素的数量。当稀疏度接近 *1* 时，矩阵高度稀疏。

例题 15.8 (稀疏矩阵示例) 以下矩阵是稀疏的：

$$\mathbf{A} = \begin{pmatrix} 5 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

总元素数：$4 \times 4 = 16$，非零元素数：*4*，稀疏度：$1 - \frac{4}{16} = 0.75$。

注 15.2 (稀疏性的量化标准) 没有绝对的阈值，但通常认为：

- 非零元素比例 $< 5\%$：稀疏矩阵
- 非零元素比例 $< 1\%$：高度稀疏矩阵
- 非零元素比例 $< 0.1\%$：极端稀疏矩阵

在科学计算中，$n \times n$ 矩阵可能只有 $O(n)$ 个非零元素，而非 $O(n^2)$。

定义 15.11 (对角矩阵) $\mathbf{D} = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ 是高度稀疏矩阵，$nnz(\mathbf{D}) = n$。

- 存储：只需存储 n 个对角线元素
- 矩阵向量乘法：$O(n)$
- 矩阵乘法：$\mathbf{DA}$ 相当于 $\mathbf{A}$ 每行缩放

定义 15.12 (三对角矩阵) 仅主对角线和相邻两条对角线非零：

$$\mathbf{T} = \begin{pmatrix} b_1 & c_1 & 0 & \cdots & 0 \\ a_2 & b_2 & c_2 & \cdots & 0 \\ 0 & a_3 & b_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & c_{n-1} \\ 0 & \cdots & 0 & a_n & b_n \end{pmatrix}$$

$nnz(\mathbf{T}) = 3n - 2$，常见于差分方程离散化。

定义 15.13 (带状矩阵) 非零元素集中在主对角线附近。带宽 w：对任意非零元素 a_{ij}，$|i - j| \leq w$。

$$nnz(\mathbf{A}) \approx n(2w + 1)$$

当 $w \ll n$ 时高效。

定义 15.14 (稀疏模式) 定义稀疏模式矩阵 $\mathbf{P} = (p_{ij})$：

$$p_{ij} = \begin{cases} 1 & \text{如果} a_{ij} \neq 0 \\ 0 & \text{如果} a_{ij} = 0 \end{cases}$$

稀疏模式在图论、符号计算中很有用。

定义 15.15 (稀疏图与社交网络) 社交网络、引文网络等可用稀疏邻接矩阵表示：

$$A_{ij} = \begin{cases} 1 & \text{用户} i \text{与用户} j \text{相连} \\ 0 & \text{否则} \end{cases}$$

通常高度稀疏（每人连接数远小于总人数）。

下面我们来看一下稀疏矩阵的数学性质：

定理 15.14 (稀疏矩阵的秩) 稀疏矩阵的秩可能远小于其维度。考虑 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 稀疏：

$$\text{rank}(\mathbf{A}) \leq \min(nnz(\mathbf{A}), n)$$

但实际中通常 $\text{rank}(\mathbf{A}) \ll n$。

定义 15.16 (稀疏矩阵的图表示) 矩阵 $\mathbf{A}$ 对应有向图 $G = (V, E)$：

- 顶点 $V = \{1, 2, \ldots, n\}$
- 边 $(i, j) \in E$ 当且仅当 $a_{ij} \neq 0$

对称矩阵对应无向图。

定理 15.15 (稀疏矩阵的谱性质) 稀疏矩阵的特征值分布：

- 可能有许多零特征值
- 非零特征值数量 $\leq nnz(\mathbf{A})$
- *Gershgorin* 圆盘定理：特征值位于圆盘 $\{z \in \mathbb{C} : |z - a_{ii}| \leq \sum_{j \neq i} |a_{ij}|\}$

15.3 Exercises